International Journal of Semantic Computing
Vol. 10, No. 2 (2016) 147–166

DOI: 10.1142/S1793351X16400055

"Climate Change" Frames Detection and Categorization Based on Generalized Concepts

Saud Alashri* and Sultan Alzahrani†
School of Computing, Informatics, and Decision Systems Engineering
Arizona State University, Tempe, Arizona, USA
**salashri@asu.edu*
†*ssalzahr@asu.edu*

Jiun-Yi Tsai‡ and Steven R. Corman§
Hugh Downs School of Human Communication
Arizona State University, Tempe, Arizona, USA
‡*jtsai8@asu.edu*
§*steve.corman@asu.edu*

Hasan Davulcu
School of Computing, Informatics, and Decision Systems Engineering
Arizona State University, Tempe, Arizona, USA
hdavulcu@asu.edu

The subliminal impact of framing of social, political and environmental issues such as climate change has been studied for a long time in political science and communications research. Media framing offers "interpretative package" for average citizens on how to make sense of climate change and its consequences to their livelihoods, how to deal with its negative impacts, and which mitigation or adaptation policies to support. A line of related work has used bag of words and word-level features to detect frames automatically in text. Such works face limitations since standard keyword based features may not generalize well to accommodate surface variations in text when different keywords are used for similar concepts. In this paper, we develop a new type of textual features that generalize (subject,verb,object) triplets extracted from text, by clustering them into high-level concepts. We utilize these concepts as features to detect frames in text. Our corpus comprises more than 45,000 climate change related sentences. Expert coders annotated those sentences as Frame/Non-Frame and framed sentences were mapped into one of four general frame categories: solution, problem threat, cause, and motivation. Compared to uni-gram and bi-gram based models, classification and clustering using our generalized concepts yielded better discriminating features and a higher accuracy classifier with a 12% boost (i.e. from 74% to 83% in f-measure) and 0.91 clustering purity for Frame/Non-Frame detection.

Keywords: Text mining; Frames Detection; Concepts; Big Data; Climate Change; Natural Language Processing; Classification; Clustering.

1. Introduction

Climate change has provoked heated debates on the global political and media arenas. Media framing offers "interpretative package" for average citizens on how to make sense of climate change and its consequences to their livelihoods, how to deal with its negative impacts, and which mitigation or adaptation policies to support [18, 38, 44]. News frames encourage salient interpretation of debated issues through the usage of rhetorical devices (e.g. words, repetitive phrases, and metaphors). Increasingly, governments and international communities are concerned about the security implications of climate change as empirical research has documented that climate change is linked to increased risk of violent conflict [6]. For example, in May 2015, U.S. President Barack Obama suggested that extreme weather is a threat to national security and elevates the risk of global instability and conflict. Some popular press adopted security threat frame to gain public attention. Therefore, systematic detection of news frames related to climate change offers better understanding of stakeholders and their competing perspectives.

Politicians have used framing on hotly debated issues to shift public opinion, gain support and pursue their agenda. A **frame** is the bundling of a component of oratory to urge certain perceptions and to dishearten others [2]. Framing is accomplished when a choice of words, expressions, subjects and other logical gadgets support one understanding of an arrangement of realities, and debilitate other interpretations. One of those framed issues is climate change. Internet created a public space for politicians and stakeholders to frame climate change and related issues to push for their agenda. Online tools such as blogsphere, microblogging and social media streams have increased the availability of data on climate change related debate and made it feasible for researchers to analyze them.

Framing research requires qualitative analysis of a number of texts by subject matter experts to identify and code a set of frames. This is a time consuming process that does not scale well. In order to address the scalability problem, machine learning techniques can be utilized to detect and classify frames. In this paper we propose a system for automatic detection of frames in sentences in a climate change related corpus, and map them to one of four expert identified frame categories: solution, problem threat, cause, and motivation. Our problem here can be described as a multi-level multi-class classification problem where we first classify each sentence as Frame or Non-Frame. Then, the Frame sentences are further mapped into one of four predefined frame categories. In particular, we show that if a sentence is (subject,verb,object) patterned generalized concepts and relations then using generalized concepts and relations [15] as features produced significant results compared to classical textual features (e.g. uni-grams and bi-grams) while detecting and categorizing Frame/Non-Frame sentences. In unsupervised learning approach, we experimented with k-means [5] and its results aligned with our development of the four frame categories using theories discussed later in Sec. 4.3. In supervised learning approach, we experimented with SVM [19], Random Forests [11] and sparse logistic regression [35] classifiers, and identified sparse logistic regression as the best performing classifier for these tasks.

Generalized concepts approach extracts high-level information from text as relationships and concepts forming a semantic network. It first uses shallow semantic parser to generate POS tags to obtain semantic triplets (subject,verb,object) from

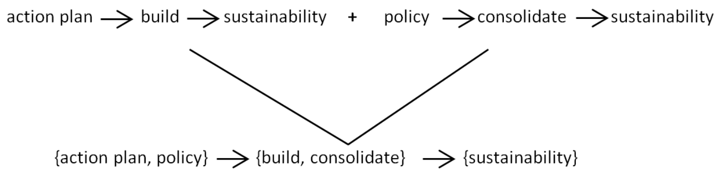

Fig. 1. Example of merging two related concepts.

text. Next, it utilizes a bottom-up agglomerative clustering approach to merge and generalize those triplets into concepts. In NLP, shallow parsing is the task of extracting the *subjects*, predicates or *verb phrases*, and *objects*. Figure 1 shows how two related triplets could be merged into a higher level generalized concept. In this figure, two extracted triplets: $\langle$*action plan* $\rightarrow$ *build* $\rightarrow$ *sustainability*$\rangle$ and $\langle$*policy* $\rightarrow$ *consolidate* $\rightarrow$ *sustainability*$\rangle$ are merged to form a high level generalized concept and relationship as: $\langle${*action plan, policy*} $\rightarrow$ {*build, consolidate*} $\rightarrow$ {*sustainability*}$\rangle$ by discovering contextual synonyms such as {*action plan, policy*} and {*build, consolidate*}. Here the definition of contextual synonyms is not based on the one in the traditional dictionary. Rather, they correspond to phrases that may occur in similar semantic roles and associated with similar contexts. In Fig. 1 the two triplets share the same object {*sustainability*} and semantically similar verbs; hence, we can merge their subjects {action plan,policy} as contextual synonyms.

2. Related Work

2.1. *Media framing*

Mainstream media serve as the main arena where international governments, social and political actors, scientists, social movement organizations interact and make competing claims about climate change issues [26]. Communication surrounding climate change can inhibit or support science and policy interactions, propagate consensus or disagreements [27], and ultimately facilitate social change [10], [37], depending on how messages about climate change have been framed [10].

Media representation of climate change plays a vital role in shaping ongoing policy discourse, public perception and attitudes. [14] suggests that prominent political actors frame climate risk for their own purposes, and align frames with their interests and perspectives through media feedback processes of representing climate change risk. Studies have shown that the lay people learn about climate change mainly through consuming mainstream media news [12]. Consequently, [38] argued news media framing can catalyze public engagement and help trigger collective concern of climate change. Put differently, media framing is a powerful tool to highlight different aspects of the policy options, and promote specific interpretations or evaluations that influence decision making [21].

Existing typologies of climate change framing, focusing on dichotomous categories, are limited by their inability to link framing processes with movement interaction. We argue that, in order to understand how the media reflect different organizations interests in addressing climate change as a social problem, it is necessary to supplement the social movement focus on resource mobilization to framing processes of collective action problems. To do that, this study develops a nuanced typology for studying climate change framing and its adequacy for supporting a social movements that would be necessary to overcome the collective action problem. Our typology provides a holistic map to evaluate how climate change media framing can enable appropriate social and policy actions that ultimately can mitigate risks of social unrest. We apply this framework to examine framing of climate change in media and social media texts collected from the Niger Basin region over seven months from August 2014 to February 2015, using a novel coding technique to assess diagnostic, prognostic, and motivational framing described by [9] as the keys to effective social movements.

2.2. *Framing research in computer science*

Jang *et al.* [28] examined twitter stream on extracted frames and pointed out a strong ties between frames collected from news with the public opinions expressed in tweeter feeds. [46] went further to distill agenda from news and link them to action. Content analysis of frames in news is performed either by (1) manual frame coding, that is done by trained coders, which is costly as well as not scalable, or by (2) frame identification by using machine learning techniques that overcome human limitations by automatically detecting frames after training a classifier [13]. Many studies have addressed media framing as a document classification problem by building a learning model to classify documents or paragraphs by utilizing different features. Aside from document level, [29], [49] examined the learning task at the sentence level and even at the phrase level. Previous work on sentence level classification has focused on experimenting with different classifiers and different features. [39] examined: bag of words, n-grams, and topic models to classify news articles and map them to a set of frames. Others, [7] employed POS-tags [47] and named entities [23] as features to detect and classify frames. Ceran *et al.* [16] experimented with {subject, verb,object} based features and benchmarked "paragraph level" classifier for story detection against standard keyword based features, which showed significant improvement in classification accuracy. More advanced conceptual features engineering was developed in [15] as they showed how generalized concepts performed better in detecting stories in paragraphs. We utilize their generalized concepts as features to detect and categorize frames. Our paper works on sentence level classification and clustering compared to their paragraph level. Also, our task is a multi-level multi-class learning task where we first examine if a sentence contains a frame, and then we identify which one of four frame categories it belongs to. Moreover, we developed triple-extraction techniques where we can extract more features and incorporate a larger percentage of sentences into the learning model (i.e. 80% of sentences compared to 40%).

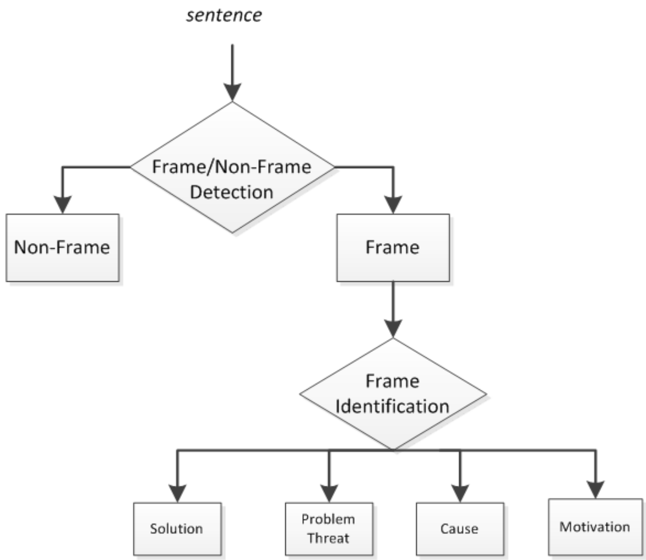

Fig. 2. Multi-level multi-class classification.

3. Problem Definition

Given a set of documents $\{D_1, \ldots, D_M\}$ where each document contains one or more paragraphs. First, we split documents into sentences $\{S_1, \ldots, S_N\}$. Next, using sentences as data points, we aim to resolve whether a sentence S_i contains a frame or not. And, if the sentence contains a frame, then we aim to identify its category, as one of: {*Solution, Problem Threat, Cause, Motivation*}. Figure 2 shows our multi-level multi-class problem for a given sentence.

4. Methodology

4.1. *Overall system model*

The overall system consists of documents collected from nearly 100 RSS feeds that are related to climate change in the Niger Delta region. We also perform sentence splitting of documents, identification of key frames and their categories by the coders, feature extraction (uni-grams, bi-grams, and generalized concepts), identification of discriminative features, and a predictive model to detect and identify the frame categories for sentences containing frame references. Figure 3 shows the main components of our system.

4.2. *Climate change corpus*

Our climate change corpus is comprised of nearly $45,054$ sentences extracted from news and social media websites, that are related to climate change topics in Niger Basin region over a seven months period from August 2014 to February 2015. There

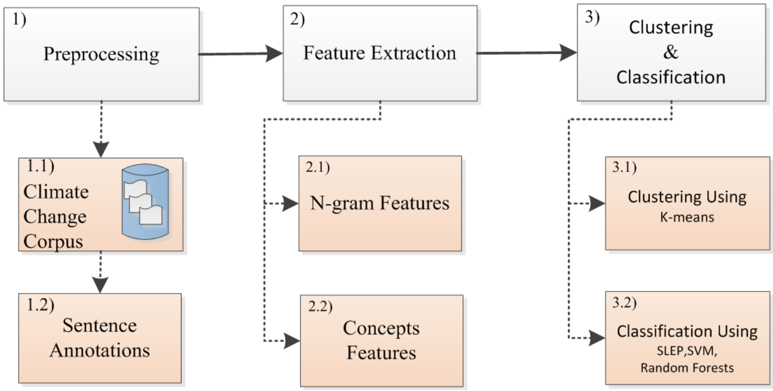

Fig. 3. System architecture.

are 16,050 sentences coded as Frame sentences and 29,004 as Non-Frame sentences by domain experts. Frame sentences are further categorized into one of four categories: Solution, Problem Threat, Cause, and Motivation.

4.3. *Development of four-class typology of media framing*

Existing typologies of climate change framing, focusing on dichotomous categories, are limited by their inability to link framing processes with movement interaction. We argue that, in order to understand how the media reflect different organizations interests in addressing climate change as a social problem, it is necessary to supplement the social movement focus on resource mobilization to framing processes of collective action problems. To do that, this study develops a nuanced typology for studying climate change framing and its adequacy for supporting a social movements that would be necessary to overcome the collective action problem. Our typology provides a holistic map to evaluate how climate change media framing can enable appropriate social and policy actions that ultimately can mitigate risks of social unrest. We apply this framework to examine framing of climate change in media and social media texts collected from the Niger Basin region, using a novel coding technique to assess diagnostic, prognostic, and motivational framing described by [9] as the keys to effective social movements.

4.3.1. *Media framing, collective action, and social movements*

In the field of social movement studies, framing has primarily been used to discuss challenges of strategy formation and implementation activists face [30]. Social movement scholars define framing as a process aimed at aligning movement meanings with the ideological perspectives of relevant audiences, including the general public, the media and policy makers [9] in order to produce action in support of ideological goals. Understanding climate change as a collective action problem makes a social movement approach to framing relevant, as framing "plays a central role in the need to mobilize resources, recognize and respond to opportunities and threats, and exercise pressure and influence by means of communication" [30]. This approach moves the study of framing beyond the limits of previous research with its focus on dichotomies, and highlights instead the potential impact of overarching framing strategies. As a complex social issue requiring engagement with multiple stakeholders and audiences (e.g. international organizations, local government, NGOs, scientists, and general public), climate change in developing countries, such as West Africa, provides fertile ground on which to explore the effectiveness of framing in propelling social movements in response to collective action problems.

Benford and Snow [9] develop a typology of social movement frames to explore signification strategies in the context of collection action. The authors assert that that the more central the framing is to the ideology of the targets of mobilization, the greater the hierarchical salience within their larger system of belief [45]. This hierarchy relies on the concept of narrative fidelity [24]: The more a frame "rings true" to the audience, the greater the salience of the frame, and the more potential it carries to influence collective action. The authors argue that "frames help render events or occurrences meaningful and thereby function to organize and guide action" [9]. This process occurs through the development, generation, elaboration, and contestation of three types of collective action frames: diagnostic, prognostic, and motivational.

The first type, diagnostic framing, seeks to remedy or alter some problematic situation or issue by identifying the source of causality, blame, and/or culpable agents [9]. The second type, prognostic framing, attempts provide a solution or plan of attack for the identified problem. While the first two functions seek to create a consensus in the audience, the third, motivational framing, is a call to action. According to [9], motivational framing attempts to engage the audience in ameliorative collective action. That is, motivational frames supply the impetus for public actions that go beyond diagnosis and prognosis, and include compelling vocabularies of severity, urgency, efficacy, and propriety [8]. To engage the public in solving social problems, organizations need to establish the severity of a particular situation, emphasize a sense of urgency of the threat, stress the likelihood of change or efficacy of taking actions, or highlight moral responsibility. This process occurs within a multi-organizational realm that includes opponents, audiences, media, bystanders, and the within the organization itself.

We argue that messages encouraging collective action are most effective when they combine these three types of frames. While Benford and Snow [9] do not address this issue, a story combining problem, solution, and motivation touches all the elements of the narrative arc [1], and is therefore more likely to be perceived as coherent [24]. Separating these elements in different messages relies on the audience to integrate them from different sources, a process vulnerable to effects of memory and involvement.

4.3.2. *A four-class typology of media framing*

Drawing from Benford and Snow [9] collective action frames for social movements, [48] developed a four-class typology of climate change framing to capture three functions diagnostic, prognostic, and motivational. As discussed earlier, those three functions of framing play an essential role in social actors' resource mobilization and participation in the political processes. Guided by Benford and Snow's framework, [48] also incorporated and modified a handful of common frames applicable to climate change identified from prior research (e.g. [38, 40]). To ensure that the four-class typology captured a full spectrum of possible frames that emerged from the West African media discourse, they further adopted an inductive approach based on a preliminary scanning of relevant texts. The final typology consisted of four framing classes and a set of twenty-five subcategories germane to climate change impacts and solutions.

Figure 4 provides an overview of four-class typology. Though Benford and Snow [9] identify three classes of frames, [48] split the diagnostic frame into two sub-classes, cause and problem/threat to capture the special diagnostic attention paid to causes in the climate change debate. Though West African discourse is likely different from Western discourse in this regard, singling out cause framing for special attention would provide maximum applicability of the four-class typology to other geographic contexts, and maintain a future basis for comparative analysis.

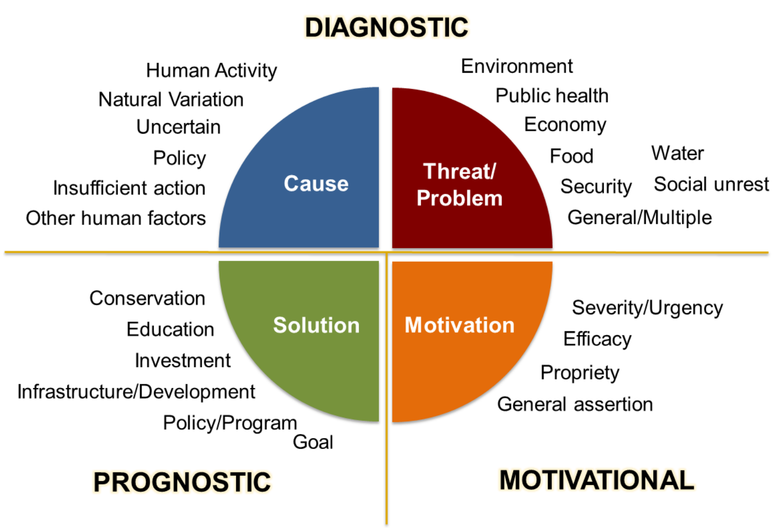

Fig. 4. Four-class typology.

4.4. *N-gram features*

We experimented with both uni-gram and bi-gram features. We run a simple term frequency-inverse document frequency (TF-IDF) [25] based technique on the entire corpus to generate a large ranked list of, stopword eliminated, uni-grams and bi-grams, and we experimented with them separately as features in our learning models.

4.5. *Generalized concepts features*

In [15], they extracted concepts from paragraphs where only 40% of the paragraphs generated concepts. In this paper, since we are working on sentence level, we improved the concept extraction approach, by extracting more triplets by utilizing a larger number of triplet extractors and pre-processing their output to include about 80% of the sentences in our experimental evaluations.

4.5.1. *Triplets extraction*

In order to extract *Subject, Verb, Object* triplets, first we run a pronoun resolver [42, 32, 33, 43]. Since triplets extraction is an ongoing research topic in NLP, we proceeded to use four state-of-the-art triplets extraction tools: ClearNLP [17], Reverb [22], Everest [3], AlchemyAPI [4] as complementary systems. Additionally, any triplet slots with phrases were segmented into keywords, stemmed, stop-word removed and their cartesian product were produced as additional triplets.

4.5.2. *Concepts generation*

Triplets extraction algorithms typically produce noisy and sparse triplets. Therefore, we apply a hierarchical bottom-up clustering algorithm that generalizes triplets into more meaningful relationships. To do so, we employ both syntactic and semantic criteria that are based on the corpus to generalize triplets into high level concepts without *drift*. In *syntactic criteria*, a pair of subjects-verbs-objects are merged only if they share common context related to their different arguments (i.e. a pair of different subjects are merged only if they co-occur with an identical verb-object context).

Additionally, we capture contextual synonyms for subjects, verbs and objects by defining a *semantic criterion* which is based on our corpus as well as WordNet [36]. Corpus-based contextual synonyms for subjects, verbs and objects is based on their common verb-object, subject-object and subject-verb contexts respectively. Also, we capture contextual synonyms that are not derivable from our corpus by applying WordNet synonyms and hyponyms on the memebers of the concepts to further expand and generalize them.

In order for the information to propagate between clusters of subjects/objects and clusters of relations, we apply a hierarchical bottom-up clustering algorithm [31]. High level concepts and relations are merged to form clusters. Each cluster is represented by graph of nodes and edges where nodes represent concepts and edges represent relations between concepts. The details of the above criteria and the generalization algorithm are available in [15].

4.6. *Unsupervised frame learning*

Unsupervised learning aims to draw inferences from given dataset where labels (i.e. classes) are hidden or unknown. It focuses on how the model can learn to represent particular input patterns in a way that reflects the statistical structure of the dataset. We utilized this approach to assist in benchmarking different features: generalized concepts, uni-grams and bi-grams in the clustering process. Our goal is to investigate which feature set will produce the best clusters. In an unsupervised learning, comparing different features sets will give a hint about the best feature set to be used in classification task. Additionally, unsupervised learning will help us in determining whether the rational and theoretical background for the development of four frame categories will align with our dataset or not. Utilizing k-means [5] we cluster the entire dataset into two clusters to see if they form Frame/Non-Frame clusters, and then the Frame sentences are clustered into four clusters mimicking the four frame categories {Solution, Problem Threat, Cause, and Motivation}.

4.7. *Supervised frame learning*

To classify each sentence as Frame/Non-Frame and identify its relevant frame category we utilize sparse learning framework [35], with the underlined motivation to select a subset of discriminating concepts that can (1) identify sentences containing frame references and (b) classify a sentence into a frame category. The following steps describe our algorithm:

1. Generate features from the entire corpus
2. Filter the features × sentences matrix to include only resultant generalized concepts/features
3. Formulate the problem in a general sparse learning framework [35]. In particular, the logistical regression formulation presented below fits this application, since it is a dichotomous frame classification problem (i.e. each sentence classified as Frame/Non-Frame), and multi-class classification problem (i.e. each Frame sentence is further classified as one of four frame {Solution, Problem Threat, Cause, and Motivation}):

$$\min_x \sum_{i=1}^{m} w_i \log(1 + \exp(-y_i(x^t a_i + c))) + \lambda|x|. \tag{1}$$

In formula (1), a_i is the vector representation of the ith sentence, w_i is the weight assigned to the ith sentence ($w_i = 1/m$ by default), and $A = [a_1, a_2, \ldots, a_m]$ is the features × sentences matrix, y_i is the label of each sentence, and the x_j, the jth element of x, is the unknown weight for each feature, ($\lambda > 0$) is a regularization parameter that controls the sparsity of the solution, $|x|_1 = \sum |x_i|$ is 1-norm of the x vector. We used the SLEP [34] sparse learning package that utilizes gradient descent approach to solve the above convex and non-smooth optimization problem. The features with non-zero values on the sparse x vector yield the discriminant factors for classifying a sentence.

5. Experimental Evaluation

5.1. *Sentence annotation*

Our experts developed four categories of climate change related frames as follows:

- **Solution framing (prognostic):** Covering the prognostic function of defining what should be done about problems, solution framing refers to actions taken to prevent further impact of climate change effects or further impact of the causes of climate change such as greenhouse gas emissions. Solutions can also emphasize ongoing measures to deal with existing effects of climate change. Six frames capture an array of mitigation and adaptation efforts conservation, education, investment, infrastructure and development, creation or implementation of policy and programs, and goal.
- **Problem Threat framing (diagnostic):** This diagnostic framing class stresses on how climate change or outcomes of climate change impact various actors, industries, human health, and the environment, Eight codes capture negative consequences and threats brought by climate change, including environmental systems and ecosystem, public health, economic development, food security, water scarcity, national security, social unrest, and general or multiple impacts. Both cause framing and problem/threat framing comprise the diagnostic function in defining social problems.
- **Cause framing:** This group of diagnostic frames focus on attributing the blame for causing climate change to either human activity, natural variation or other reasons. Six subcategories captured different explanations for causal attribution of climate change: (a) human activity, (b) natural variation, (c) scientific uncertainty, (d) policy causes, (e) insufficient actions, and (f) human disruption to mitigate climate change impact.
- **Motivation framing (motivational):** Motivational framing refers to statements that explicit call for definitive course(s) of action and explain why the audience should make an effort to enact solutions [9]. In other words, motivational frames elaborate on the rationale for action that goes beyond diagnosis and prognosis, and include vocabularies of severity, urgency, efficacy, and propriety [8]. We added a general category to analyze statements that call for actions without providing readers with above-mentioned reasons.

We assigned sentence annotation to three different expert coders where we break ties by using the majority vote.

5.2. *Quantitaive evaluation*

5.2.1. *Unsupervised learning*

Experimenting with unsupervised learning reveals dataset structure and can infer relations among data points. In this experiment, we ignored labels and clustered our dataset using three sets of features (i.e. uni-gram keywords, bi-gram terms, and

Table 1. Clustering into two clusters.

Method	SSE	Purity
Concepts	**54,322.08**	**0.91**
Bi-grams	720,044.21	0.71
Uni-grams	306,124.03	0.68

Table 2. Clustering into four clusters.

Method	SSE	Purity
Concepts	**34,397.75**	**0.98**
Bi-grams	139,124.43	0.91
Uni-grams	292,812.30	0.51

generalized concepts) separately as features and the k-means [5] as clustering algorithm. We experimented with k = 2 for the entire dataset, and k = 4 for the Frame sentences. To evaluate k-means clustering results, we utilized SSE (sum of squared error) as well as purity.

Table 1 shows the SSE and purity for clustering the entire dataset into two clusters using different features. Using generalized concepts as features, the resultant SSE (54,322.08) and purity (0.91) outperform those ones with uni-grams SSE (306,124.03) and purity (0.68) as well as bi-grams SSE (720,044.21) and purity (0.71).

Table 2 presents the SSE and purity for clustering the frame sentences into four clusters using different features. Using generalized concepts as features, the resultant SSE (34,397.75) and purity (0.98) outperform those ones with uni-grams SSE (292,812.30) and purity (0.51) as well as bi-grams SSE (139,124.43) and purity (0.91).

Figure 5 shows the resultant clusters of frame sentences using concepts as features. In this figure we have the ground truth (i.e. which sentence belongs to which

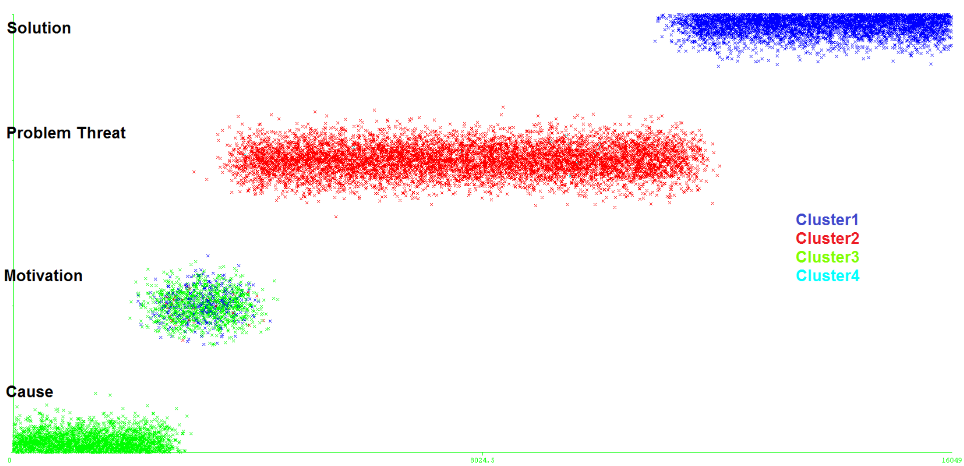

Fig. 5. Clustering frame sentences into four clusters using concepts.

frame category) by using sentence id in x-axis and the corresponding frame category in y-axis. The three clusters (1,2,3) corresponding to frames (Solution, Problem threat, and Cause) are well clustered in terms of purity. The Motivation frame (cluster 4 next to Motivation) is mixed of the other three clusters (1,2,3). Our interpretation for this impurity is that in motivational framing, typically people show the cause of a problem and propose a solution. As a result, a sentence belongs to motivation frame category could carry other frame categories (Solution, Problem threat, and Cause).

Figure 6 shows the purity for experimenting different K values (i.e. number of clusters), using concepts as features on the Frame sentences. In this figure, the highest purity is 0.98 when K = 4 which aligns with the development of the four frame categories {Solution, Problem Threat, Cause, and Motivation} discussed in Sec. 4.3.

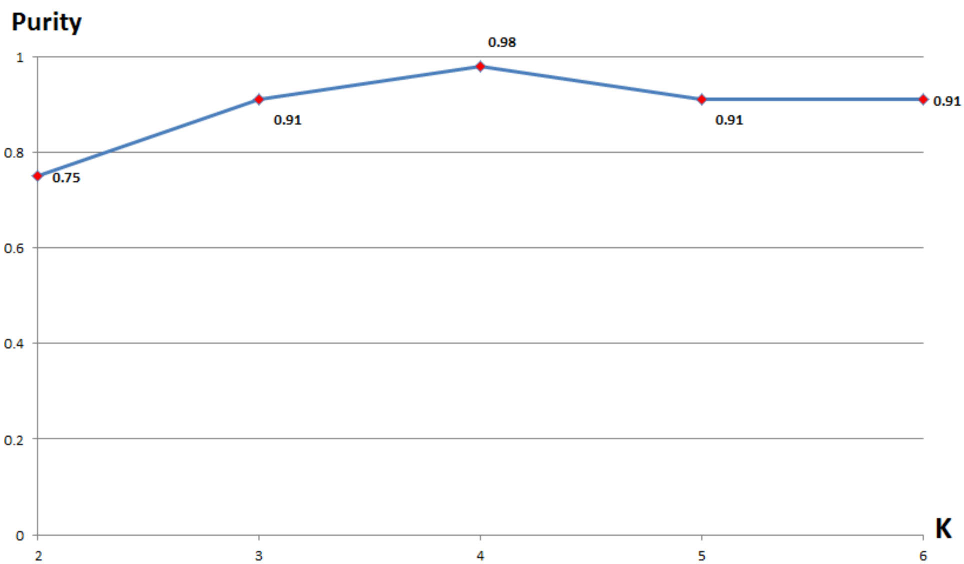

Fig. 6. Experimenting different values of K.

5.2.2. *Supervised learning*

In this approach, we use labeled dataset. Once sentences are labeled as Frame/Non-Frame and categorized with their corresponding frame category, we utilize uni-gram keywords, bi-gram terms, and generalized concepts separately as features and the sparse logistical regression classifier SLEP [34] to identify weighted discriminative features and classify sentences. We experimented with three different classifiers (SVM [19], Random Forests [11]) and found that SLEP outperformed both these other classifiers. Using different types of features generated from the entire corpus, we perform ten-fold cross-validation for measuring the classifier's predictive accuracy

Table 3. Frame/non-frame classification.

Method	Class label	Precision	Recall	F-measure
Concepts	Frame	0.80	0.88	0.84
	Non Frame	0.87	0.77	0.82
	Average	**0.83**	**0.83**	**0.83**
Bi-grams	Frame	0.75	0.42	0.54
	Non Frame	0.74	0.92	0.82
	Average	**0.74**	**0.67**	**0.68**
Uni-grams	Frame	0. 75	0.48	0.59
	Non Frame	0.76	0.91	0.89
	Average	**0.75**	**0.70**	**0.74**

to detect Frame/Non-Frame sentences. Next, using features generated from frame sentences only, we train a multi-class model to classify sentences into their corresponding frame category. We report precision, recall, and F-measure as quantitative evaluation metrics. Qualitative analysis of the identified discriminating concepts is also presented in the next section.

Table 3 presents the accuracies for detecting Frame/Non-Frame sentences using different features. Using generalized concepts approach as features, the resultant average accuracy (F-measure of 83%) outperforms both accuracies with uni-grams (74%) and bi-grams (68%) features by 12% and 22% respectively.

Table 4 (next page) shows the accuracies for identifying the corresponding frame category. Using generalized concepts, these accuracies vary between 73% and 83% (F-measure) for different categories. In this table, utilizing generalized concepts yields slightly better performance compared to both uni-grams and bi-grams with an overall average accuracy (F-measure) of 79%.

Table 4. Frame classification into four categories.

Method	Frame category	Precision	Recall	F-measure
Concepts	Solution	0.75	0.93	0.83
	Problem Threat	0.77	0.84	0.79
	Cause	0.85	0.77	0.80
	Motivation	0.89	0.62	0.73
	Average	**0.82**	**0.79**	**0.79**
Bi-grams	Solution	0.87	0.77	0.81
	Problem Threat	0.84	0.77	0.80
	Cause	0.86	0.73	0.76
	Motivation	0.90	0.58	0.71
	Average	**0.87**	**0.71**	**0.77**
Uni-grams	Solution	0.78	0.87	0.82
	Problem Threat	0.81	0.81	0.81
	Cause	0.83	0.62	0.82
	Motivation	0.85	0.57	0.64
	Average	**0.82**	**0.72**	**0.77**

5.3. *Qualitative analysis of resultant concepts*

Table 5 shows top five discriminative concepts for each frame category. Our team of experts explored the highly significant generalized concepts germane to four-class framing in media discourse surrounding climate change across West African RSS feeds and provided qualitative evaluations as follows:

Table 5. Top five generated concepts for each frame category.

Cause	Problem threat	Solution	Motivation
{Greenhouse,Emissions, Gases} ↓ {Cause,Attribute to} ↓ {Global warming}	{Flood} ↓ {Associate,Create} ↓ {Poverty,Disease}	{Action plan,Policy} ↓ {Build,Consolidate} ↓ {Sustainability,Re-silience future}	{International, Community} ↓ {Urge,Warn} ↓ {Threat}
{Industry,Anthro-pogenic} ↓ {Raise} ↓ {Earth temperature, CO2,CO5}	{Heavy rainfall, Torren-tial rain} ↓ {Create,Bring, Increase} ↓ {Flooding,Disaster, Landslide}	{Development, Sustain-ability,National program} ↓ {Enhance} ↓ {Community}	{Agreement,Leaders, World} ↓ {Help} ↓ {Future,Hope}
{Fossil fuel} ↓ {Impact,Harm} ↓ {Planet,Environment, Weather}	{Drought} ↓ {Cause,Impact, Reduce} ↓ {Food-shortage,Food-production,Crop}	{Brown} ↓ {Sign} ↓ {Local legislation, CA groundwater, Management framework}	{USA,EU,China} ↓ {Recognize,Reduce} ↓ {Emissions}
{Coal combustion,Diesel, Man-Made} ↓ {Create} ↓ {Extreme weather, Temperature-up}	{Sea-level rise} ↓ {Result in,Cause} ↓ {Tsunami,Damage, Flood}	{Sustainability,Energy} ↓ {Can help,Improve} ↓ {Food security, Households}	{Africa} ↓ {Need,Implement} ↓ {Policy,Awareness, Partnership}
{Truck,Car} ↓ {Rise} ↓ {Carbon pollution, Pollute}	{Extreme Weather, Hailstorm} ↓ {Cause,Affect} ↓ {Mudslide,Floods, Farming}	{Smart agriculture, Africa countries} ↓ {Meet,Breathe} ↓ {Life}	{Nigerian} ↓ {Apply,Take} ↓ {Measures,Renew-able Energy,Policy}

5.3.1. *Cause framing*

Causal responsibility of climate change and its effects was often attributed to anthropogenic activities, particularly man-made greenhouse gas emissions, human-induced pollution, and fossil fuel use. Carbon dioxide and greenhouse gas emission emerged as highly significant concepts, as indicated by high weigh value. Media texts often associated global warming with carbon dioxide emissions using the following triplets to construct a cohesive story:

- Scientific research indicate that atmospheric carbon dioxide increase at a large level.
- Cars and trucks were major sources of air pollution and carbon dioxide emissions, which directly increased local temperature.

5.3.2. *Problem threat framing*

Next, we turned our attention to identify the dominant concepts representing the problem and threat framing of climate change. Media texts tended to highlight devastating environmental impacts caused by climate change, such as floods, prolonged drought, loss of landmass and soil, desertification, sea-level rise, storm surge, heat waves, and more. Flooding, in particular, is a severe concern as nine out of sixteen triplets of high weigh values explicitly mentioned the negative impacts of heavy rainfall or torrential rain. Consequently, economic condition and food insecurity were influenced, infrastructure was damaged, and health diseases were exacerbated with the increased intensity and frequency of floods.

5.3.3. *Solution framing*

The most representative discourse of solution framing is discussed next in Sec. 5.3.4.

5.3.4. *Motivation framing*

When discussing motivation for why policy actors and citizens should act upon, the most salient concepts emphasized that international communities (e.g. U.S., EU, and China) should negotiate a legal agreement to reduce greenhouse gas emissions at the end of 2015. There is little attention to stating specific reasons for offering localized adaptation strategies that people can undertake. Although the awareness of climate change impacts among African government officials was generally high, the prevailing generalized concept of calling for international actions on mitigation from mainstream media discourse reflected a lack of effective national and local polices.

5.4. ***Visualizing concepts***

To visualize the generalized concept and relation clusters, we utilize a semantic network [41] of nodes (V) and edges (E) to describe the semantic space of the underlying texts. Circle nodes represent *subjects/objects* and square nodes represent *verbs*. Edges represent relations between concepts. In such a network, distinct combinations of actors (subjects) perform or recommend various sets of actions (verbs) on distinct combinations of targets (objects). The sample semantic network in Fig. 7 illustrates how *sustainability* emerges as a concept that is central to

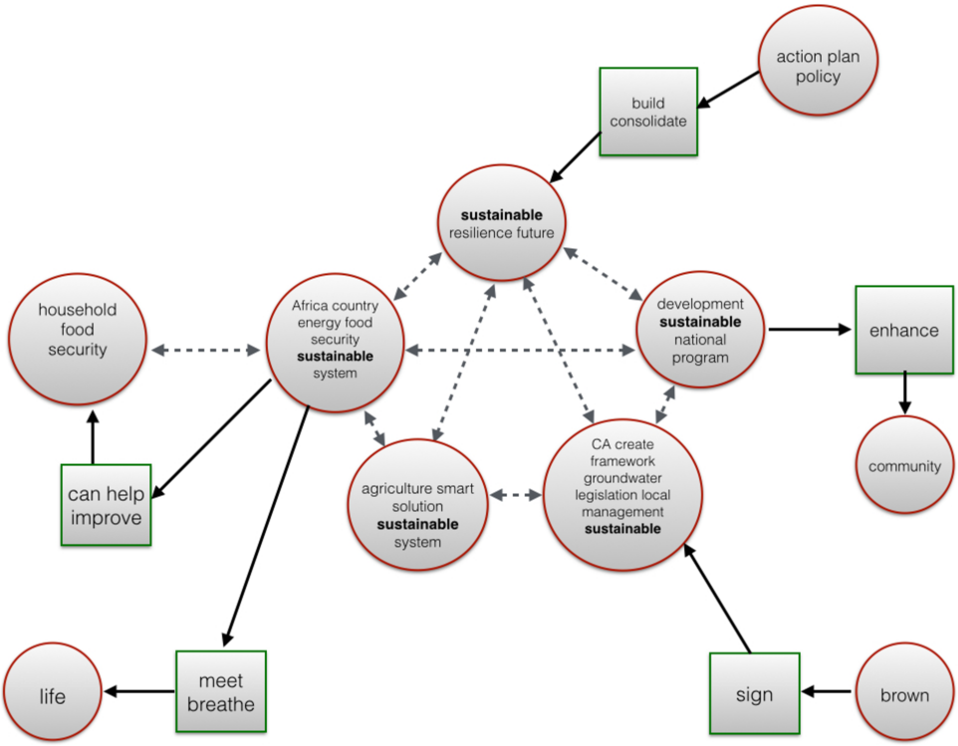

Fig. 7. A sample semantic network of frame concepts.

addressing climate change impacts. The semantic network represents the contextual relationships between generalized triplets relating to strategies for sustainable adaptation. In the media discourse, sustainable adaptation is predominantly framed as an effective solution to reduce impacts of climate change and contribute to social, economic, and environmental development. As shown in Fig. 7, developing sustainable national programs (or actions) can enhance local community resilience. According to the IPCC (Intergovernmental Panel on Climate Change) report, majority of rural communities rely on rain-fed agriculture to sustain their livelihoods in West Africa, the region worst affected by climate change. With changing rainfall patterns, prolonged droughts and flooding, sustainable system of developing agriculture-smart technologies can help improve food security at the household level. Interestingly, the African media discussed that California Governor Jerry Brown has signed the most significant framework for regulating underground water resources to achieve sustainable development in September, 2014.

6. Conclusion and Future Work

Climate change framing has pervasive influence, and this paper presents a new computational approach based on generalized concepts to identify popular media frames and map them to different categories: solution, problem threat, cause, and motivation. A line of related work has used bag of words and word-level features to detect frames automatically in text. Such work face limitations since standard keyword based features may not generalize well to accommodate surface variations in text when different keywords are used for similar concepts. In this paper, we developed a new type of textual features that generalize (subject,verb,object) triplets extracted from text, by clustering them into high-level concepts. Compared to unigram and bi-gram based models, frame classification and clustering using our generalized concepts yielded better discriminating features with a 12% boost in accuracy (i.e. from 74% to 83% in f-measure) and 0.91 clustering purity for Frame/Non-Frame detection. In our future work, we plan to utilize discriminating generalized concepts indicating actor-action-target sequences to infer causal chains of events, frames, and actions that might lead to better indicators of climate-change related social unrest.

Acknowledgments

Some of the material presented here was sponsored by Department of Defense and it is approved for public release, case number: 15–467.

References

[1] H. Porter Abbott, *The Cambridge Introduction to Narrative* (Cambridge University Press, 2008).

[2] Saud Alashri, Sultan Alzahrani, Lenka Bustikova, David Siroky, and Hasan Davulcu, What animates political debates? Analyzing ideological perspectives in online debates between opposing parties, in *Proceedings of the ASE/IEEE International Conference on Social Computing (SocialCom-15)*, 2015.

[3] Everest triplet extraction, Next Century Corporation (2013) [Online]. Available: https://github.com/NextCenturyCorporation/EVEREST-TripletExtraction

[4] Alchemyapi language features, AlchemyAPI, Inc. (2015) [Online]. Available: http://www.alchemyapi.com/products/alchemylanguage

[5] D. Arthur and S. Vassilvitskii, k-means++: the advantages of careful seeding, in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2007, pp. 1027–1035.

[6] Jon Barnett and W. Neil Adger. Climate change, human security and violent conflict, *Political Geography* **26**(6) (2007) 639–655.

[7] E. Baumer, E. Elovic, Y. Qin, F. Polletta and G. Gay, Testing and comparing computational approaches for identifying the language of framing in political news, in *HLT-NAACL*, 2015.

[8] Robert D. Benford, "You could be the hundredth monkey": Collective action frames and vocabularies of motive within the nuclear disarmament movement, *Sociological Quarterly*, 1993, pp. 195–216.

[9] Robert D. Benford and David A. Snow, Framing processes and social movements: An overview and assessment, *Annual Review of Sociology*, 2000, pp. 611–639.

[10] Maxwell T. Boykoff, *Who Speaks for the Climate?: Making Sense of Media Reporting on Climate Change* (Cambridge University Press, 2011).

[11] Leo Breiman, Random forests, *Mach. Learn.* **45**(1) (2001) 5–32.

[12] Robert J. Brulle, Jason Carmichael and J. Craig Jenkins, Shifting public opinion on climate change: An empirical assessment of factors influencing concern over climate change in the US, 2002–2010, *Climatic Change* **114**(2) (2012) 169–188.
[13] Björn Burscher, Daan Odijk, Rens Vliegenthart, Maarten de Rijke and Claes H. de Vreese, Teaching the computer to code frames in news: Comparing two supervised machine learning approaches to frame analysis, *Communication Methods and Measures* **8**(3) (2014) 190–206.
[14] Anabela Carvalho, Ideological cultures and media discourses on scientific knowledge: Re-reading news on climate change, *Public Understanding of Science* **16**(2) (2007) 223–243.
[15] B. Ceran, N. Kedia, S. Corman and H. Davulcu, Story detection using generalized concepts and relations, in *Proceedings of International Symposium on Foundation of Open Source Intelligence and Security Informatics, in conjuction with IEEE ASONAM*, 2015.
[16] Betul Ceran, Ravi Karad, Ajay Mandvekar, Steven R. Corman and Hasan Davulcu, A semantic triplet based story classifier, in *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2012, pp. 573–580.
[17] Jinho D. Choi, *Optimization of Natural Language Processing Components for Robustness and Scalability*, PhD thesis, University of Colorado at Boulder, 2012.
[18] Dennis Chong and James N. Druckman, A theory of framing and opinion formation in competitive elite environments, *Journal of Communication* **57**(1) (2007) 99–118.
[19] Corinna Cortes and Vladimir Vapnik, Support-vector networks, *Machine Learning* **20** (3) (1995) 273–297.
[20] Mary Douglas, *Natural Symbols: Explorations in Cosmology* (Routledge, 2004).
[21] Robert M. Entman, Framing: Towards clarification of a fractured paradigm, *McQuail's Reader in Mass Communication Theory*, 1993, pp. 390–397.
[22] Anthony Fader, Stephen Soderland and Oren Etzioni, Identifying relations for open information extraction, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, Scotland, UK, July 27–31, 2011.
[23] J. R. Finkel, T. Grenager and C. Manning, Incorporating non-local information into information extraction systems by gibbs sampling, in *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, 2005, pp. 363–370.
[24] Walter R. Fisher, Narration as a human communication paradigm: The case of public moral argument, *Communications Monographs* **51**(1) (1984) 1–22.
[25] J. A. Hartigan and M. A. Wong Algorithm AS 136: A k-means clustering algorithm, *Applied Statistics* **28**(1) (1979) 100–108.
[26] Stephen Hilgartner and Charles L. Bosk, The rise and fall of social problems: A public arenas model, *American Journal of Sociology*, 1988, pp. 53–78.
[27] Mike Hulme, *Why We Disagree about Climate Change: Understanding Controversy, Inaction and Opportunity* (Cambridge University Press, 2009).
[28] S. Mo Jang and P. Sol Hart, Polarized frames on climate change and global warming across countries and states: Evidence from twitter big data, *Global Environmental Change*, **32** (2015) 11–17.
[29] Yoon Kim, Convolutional neural networks for sentence classification, *arXiv preprint arXiv:1408.5882*, 2014.
[30] Graham Knight and Josh Greenberg, Talk of the enemy: Adversarial framing and climate change discourse, *Social Movement Studies* **10**(4) (2011) 323–340.
[31] Stanley Kok and Pedro Domingos, Extracting semantic networks from text via relational clustering, in *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases — Part I*, 2008, pp. 624–639.

[32] H. Lee, Y. Peirsman, A. Chang, N. Chambers, M. Surdeanu and D. Jurafsky, Stanford's multi-pass sieve coreference resolution system at the CoNLL-2011 shared task, in *CoNLL 2011*, 2011, p. 28.
[33] Heeyoung Lee, Angel Chang, Yves Peirsman, Nathanael Chambers, Mihai Surdeanu and Dan Jurafsky, Deterministic coreference resolution based on entity-centric, precision-ranked rules. *Comput. Linguist.* **39**(4) (2013) 885–916.
[34] J. Liu, S. Ji and J. Ye, *SLEP: Sparse Learning with Efficient Projections*, Arizona State University, 2009.
[35] Jun Liu, Jianhui Chen and Jieping Ye, Large-scale sparse logistic regression, in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009, pp. 547–556.
[36] George A. Miller, Wordnet: A lexical database for English, *Commun. ACM* **38**(11) (1995) 39–41.
[37] Susanne C. Moser and Lisa Dilling, *Creating a Climate for Change* (Cambridge University Press, 2006).
[38] Matthew C. Nisbet, Communicating climate change: Why frames matter for public engagement, *Environment: Science and Policy for Sustainable Development* **51**(2) (2009) 12–23.
[39] Daan Odijk, Björn Burscher, Rens Vliegenthart and Maarten De Rijke, Automatic thematic content analysis: Finding frames in news, in *Social Informatics*, 2013, pp. 333–345.
[40] Saffron O. Neill, Hywel T. P. Williams, Tim Kurz, Bouke Wiersma, and Maxwell Boykoff, Dominant frames in legacy and social media coverage of the IPCC Fifth Assessment Report, *Nature Climate Change* **5**(4) (2015) 380–385.
[41] M. R. Quillian, Semantic memory, in *Semantic Information Processing*, 1968, pp. 227–270.
[42] K. Raghunathan, H. Lee, S. Rangarajan, N. Chambers, M. Surdeanu, D. Jurafsky and C. Manning, A multi-pass sieve for coreference resolution, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2010, pp. 492–501.
[43] Marta Recasens, Marie C. de Marneffe and Christopher Potts, The life and death of discourse entities: Identifying singleton mentions, in *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2013, pp. 627–633.
[44] Adam Shehata and David Nicolas Hopmann, Framing climate change, *Journalism Studies* **13**(2) (2012) 175–192.
[45] David A. Snow and Robert D. Benford, Master frames and cycles of protest, *Frontiers in Social Movement Theory*, 1992, pp. 133–155.
[46] Katsiaryna Stalpouskaya and Christian Baden. To do or not to do: The role of agendas for action in analyzing news coverage of violent conflict, *ACL-IJCNLP 2015*, 2015, p. 21.
[47] Kristina Toutanova, Dan Klein, Christopher D. Manning and Yoram Singer, Feature-rich part-of-speech tagging with a cyclic dependency network, in *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Vol. 1*, 2003, pp. 173–180.
[48] Jenny Tsai, Steven R. Corman, M. Nolen and K. Fleischer, Toward a nuanced typology of media discourse of climate change, impact, and adaptation: An analysis of West African online news and social media, in *The Annual Convention of the Association for Education in Journalism and Mass Communication*, 2015.
[49] Theresa Wilson, Janyce Wiebe and Paul Hoffmann, Recognizing contextual polarity in phrase-level sentiment analysis, in *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 2005, pp. 347–354.