Stretching the Life of Twitter Classifiers with Time-Stamped Semantic Graphs

A. Elizabeth Cano,[1] Yulan He,[2] and Harith Alani[1]

[1] Knowledge Media Institute, Open University, UK
ampaeli@gmail.com, h.alani@open.ac.uk
[2] School of Engineering and Applied Science, Aston University, UK
y.he@cantab.net

Abstract. Social media has become an effective channel for communicating both trends and public opinion on current events. However the automatic topic classification of social media content pose various challenges. Topic classification is a common technique used for automatically capturing themes that emerge from social media streams. However, such techniques are sensitive to the evolution of topics when new event-dependent vocabularies start to emerge (e.g., Crimea becoming relevant to War_Conflict during the Ukraine crisis in 2014). Therefore, traditional supervised classification methods which rely on labelled data could rapidly become outdated. In this paper we propose a novel transfer learning approach to address the classification task of new data when the only available labelled data belong to a previous epoch. This approach relies on the incorporation of knowledge from DBpedia graphs. Our findings show promising results in understanding how features age, and how semantic features can support the evolution of topic classifiers.

Keywords: social media, topic detection, DBpedia, concept drift, feature relevance decay

1 Introduction

Microbloging platforms such as Twitter, has proven to be powerful tools for sharing opinions and spreading the word on trends and current events. Understanding what is being discussed on social media has been the focus of much research and development, to monitor opinion and sentiment [21, 11], to detect emerging events [27, 8], to track topics [5, 12], etc. One persistent challenge often faced by such works is the task of assigning topic labels to microposts; a core step in classifier training. The continuous change in topics and vocabulary on social media raises the need for retraining such classifiers with fresh topic-label annotations, which are often time consuming and costly to acquire. Topic classification of microposts is also challenged by the inherent characteristics of social media content, which often consists of ill-formed language, abbreviations, and hashtags.

In an event-dependent topic, not only new lexical features could potentially recharacterise the topic, but also previous features could fade out and become irrelevant for this topic. Because of the progressive feature drifts of topics in dynamic environments the expectation that training data and future data to be in the same feature space is not normally met. One such topic is Violence in Social Media and microposts, whose language model is continuously reshaping based on current violence-related events. For example, the word Crimea might not have been relevant to the topic Violence two years ago, but has become increasingly relevant in recent months. Similarly, the term Jan25, which was characteristic of violence behaviour during the Egyptian revolution, is now less representative of violence in current microblogs.

Such concept drifts [9][16] introduce new challenges to the topic classification of tweets. These linguistic and topic evolutions contribute to the progressive reshaping of the language model that characterises a topic, which renders existing topic classification models less and less efficient. To maintain the adequacy of our models, it is necessary to regularly retune them to fit current social media content. Relearning the models would enable us to incorporate new relevant features, and to reuse the weight of features which have become outdated or less relevant to the topic.

Particularly on a topic classification of tweets at a current epoch, it is common to only have sufficient training data from previous epochs. An extensive area of research which addresses this problem is Transfer Learning [18], which aims to apply knowledge learned in the past to solve new problems.

In this paper we propose a transfer learning approach to the epoch-based topic classification of tweets, where no label data is available on a current epoch but label data from past epochs is available. This approach relies on the incorporation of semantic features derived from temporal topic graphs extracted from a structured knowledge source. DBpedia has become one of the major sources of structured knowledge extracted from Wikipedia. Such structures gradually re-shape the representation of Topics as new events relevant to such topics emerge. The incorporation of new event-data to a topic representation leads to a linguistic evolution of a topic, but also to a change on its semantic structure. To the best of our knowledge, none of the existing approaches for topic classification using semantic features [10][3][26], has focused on the epoch-based transfer learning task. In this work we present a comparison of lexical and semantic features on epoch-based transfer learning tasks. The main contributions of this paper can be summarised as follows:

(1) we generate a cross-epoch dataset consisting of 12,000 annotated tweets over three different years and three topics;
(2) we enrich our classification models with 4 types of semantic features extracted from our Twitter content using different DBpedia dumps (3.6 to 3.9) to simulate epoch-based settings;
(3) we propose a novel weighting strategies for epoch-based transfer learning which relies on topic-based semantic graphs at a given point in time. Our findings show that the proposed strategies improve performance upon our baseline while outperforming F-measure upon lexical features; and
(4) we compare the performance of lexical feature-based models against semantic features. Our findings demonstrate that class-based (rdf:type) features alone can achieve on average a gain in F of 12% over lexical features on cross-epoch settings.

2 Related Work

Topic classification of tweets consists of the task of labelling a tweet as being either *topic-related* or *non-topic-related*. Various works have made use of lexical and profile based features to approach this task [19, 24]. Other approaches have incorporated the use of external knowledge sources (KS) to enrich Twitter content. Some of them relying only on KS [10, 23, 17]; others incorporating semantic features derived from semantic meta graphs [26, 3] on supervised settings; and others incorporating DBpedia lexical features on unsupervised classification tasks [2]. However to the best of our knowledge, none of these approaches focused on the epoch-based transfer learning task. In contrast to previous work, rather than focusing on how semantic features perform against lexical features within the same epoch datasets, we focus on analysing the change in performance on cross-epoch settings. In these settings, models are trained on data from an epoch t, and tested on data for which no training data is available yet.

Transfer learning was proposed over a decade ago [25, 4]. However, its use in natural language processing is relatively new[18]. [1] introduced a structural correspondence learning method for domain adaptation applied to part-of-speech tagging. [7] introduced the feature augmentation strategy for domain adaptation. [15] studied cross-domain classification by applying word similarities using semantic nets. However, their setting is not cross-epoch dependent but rather cross-domain. Previous work on sentiment analysis [12] studied the simultaneous sentiment and topic detection on a dynamic setting based on an unsupervised approach. As opposed to previous work which rely on the use of lexical features, we propose the incorporation of semantic features in the cross-epoch learning task. To the best of our knowledge no existing work has been formally studied for the topic classification of tweets as a cross-epoch transfer learning task on a supervised setting.

3 Characterising Topic Changes with DBpedia

DBpedia is periodically updated to incorporate any additions and modification in Wikipedia. This enables us to track how specific resources evolve over time, by comparing these resources over subsequent DBpedia editions.

For example, changes to the semantic graph for the concept Barack_Obama can be derived from snapshots of this resource's semantic graph from different DBpedia dumps.[3]. Consider Figure 1, although some of the triples remain unchanged in consecutive dumps, (e.g. `[dbp:Barack_Obama, dbo:birthPlace, dbpedia:Hawaii]`) new triples provide further information on the resource: i) current contexts (e.g. `DBpedia 3.7 [dbp:Barack_Obama, skos:subject, dbp:Al-Qaeda]`); ii) future contexts (e.g. `DBpedia 3.7 [dbp:Barack_Obama, dbo:wikiPageWikiLink, dbp:United_States_presidential_candidates,_2012]`) and iii) past context (e.g. `DBpedia 3.8 [dbp:Barack_Obama, dbo:wikiPageWikiLink, dbp:Budget_Control_Act_of_2011]`). Changes regarding a resource are exposed both through new semantic features (i.e triples) and new lexical features –appearing on changes in a resource's abstract–.

DBpedia therefore covers a wealth of structured resources exhibiting both lexical and semantic information. Moreover, these resources are commonly characterised with a Topic via the `skos:subject` property, which links a DBpedia resource with a `skos:Concept`. Hence in DBpedia each particular topic (e.g. `cat:War`[4]) is broadly represented through its associations with a large number of resources (e.g. `dbp:War_profiteering`). This resource-concept relationship yields to a broad set of resources characterising a topic. A topic can be therefore represented by a collection of resources belonging to both the main topic (e.g. `cat:War`) and resources (e.g `dbp:Combat_assessment`) belonging to subcategories (e.g. `cat:Military_operations`) of the main Topic.

Using multiple DBpedia dumps, we are able to characterise topics during different time periods. This paper proposes a novel approach which makes use of time-based semantic graph changes for characterising the relevance of a feature to a given Topic. The

[3] The DBpedia dumps correspond to Wikipedia articles at different time periods as follows: DBpedia 3.6 generated on 2010-10-11; DBpedia 3.7 on 2011-07-22, DBpedia 3.8 on 2012-06-01, DBpedia 3.9 on late April. DBpedia have them available to download at DBpedia http://wiki.dbpedia.org/Downloads39

[4] Where cat is the qname for http://dbpedia.org/resource/Category:

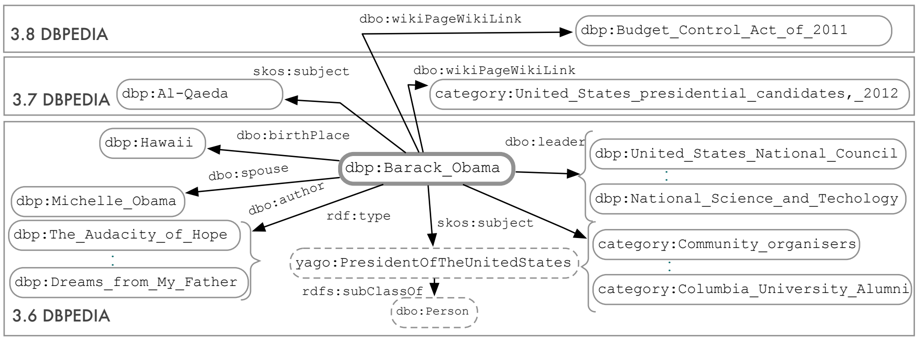

Fig. 1. Triples of the Barack_Obama resource extracted from different DBpedia dumps (3.6 to 3.8). Each DBpedia dump presents a snapshot in time of factual information of a resource.

following section introduces our framework for extracting a time-dependent DBpedia-based representations of tweets. It also presents a set of feature weighting strategies which aim to overcome the drop in classification performance when classifiers are applied to previously unseen datasets.

4 Framework for Twitter Topic Classification with DBpedia

Since the changes on the lexical and semantic representation of a topic are time-dependent, we propose to make use of temporal features in the form of semantic-graphs snapshots. In this paper we aim to understand how the relevance of features for classifying a topic changes once the characterisation of that topic changes over time. To this end, we perform an analysis based on the lexical and semantic feature expansion of tweets using DBpedia[5]. This involves investigating how the availability of resources overtime can impact the classification performance on previously unseen data. As depicted in Figure 2, our framework makes use of different DBpedia dumps for the topic classification of tweets. The main stages of this framework are: 1) Extraction of lexical and semantic features from tweets; 2) Time-dependent content modelling; 3) Strategy for weighting topic-relevant features with DBpedia; and 4) Construction of time-dependent topic classifiers based on lexical, semantic and joint features . These stages are described in the following subsections.

4.1 Lexical and Semantic Feature Extraction

We focus on two main feature types: lexical and semantic features. The lexical feature representation of a tweet consists of a bag of words approach using a TF-IDF weighting strategy [13]. To generate a semantic feature representation of a tweet, we make use of DBpedia information for all entities appearing on this content. The semantic feature generation consists of three stages: 1) entity extraction; 2) entity linking to DBpedia resources, and 3) generation of semantic features. We first extract entities from a tweet content using the AlchemyAPI entity extraction and Linked Data service.[6] This service takes a piece of text as an input, and returns a collection of annotated entities appearing in the given text. Each entity annotation provides both the entity type and a set of

[5] Analysis of joint KSs is future work

[6] AlchemyAPI, `http://www.alchemyapi.com/`

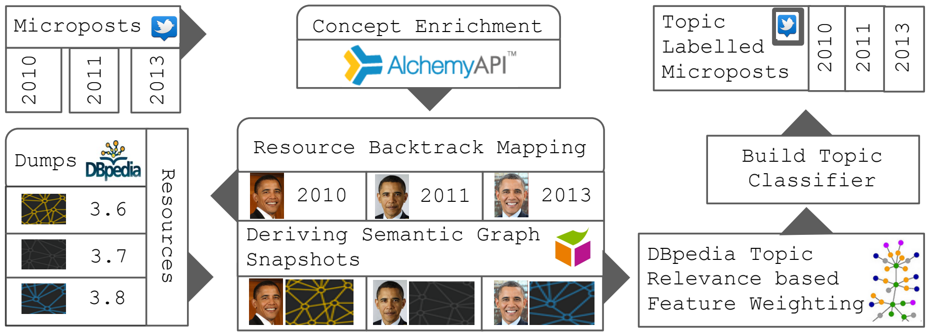

Fig. 2. Architecture for backtrack mapping of resources to DBpedia dumps and deriving topic-relevance based features for epoch-dependent topic classification.

disambiguated links for this entity. An entity's disambiguated links include links pointing to DBpedia, Freebase,[7] and Yago [8] resources. In this analysis we only kept entities disambiguated to DBpedia resources. The following section describes the generation of time-based semantic features.

4.2 Time-based Content Modeling

A *Resource Meta Graph* is an aggregation of all resources, properties and classes related to a resource [3]. Here we extend this definition by assigning a temporal marker to this graph:

Definition 1 (Resource Meta Graph) *is a sequence of tuples* $\mathtt{G} := (R, P, C, Y, f_t)$ *where*

- $\mathtt{R}$, $\mathtt{P}$, $\mathtt{C}$ *are finite sets whose elements are* `resources`, `properties`, *and* `classes`*;*
- $\mathtt{Y}$ *is the ternary relation* $\mathtt{Y} \subseteq \mathtt{R} \times \mathtt{P} \times \mathtt{C}$ *representing a hypergraph with ternary edges. The hypergraph of a Resource Meta Graph* $\mathtt{Y}$ *is defined as a tripartite graph* $\mathtt{H(Y)} = \langle \mathtt{V}, \mathtt{D} \rangle$ *where the vertices are* $\mathtt{V} = \mathtt{R} \cup \mathtt{P} \cup \mathtt{C}$*, and the edges are:* $\mathtt{D} = \{\{r, p, c\} \mid (r, p, c) \in Y\}$.
- f_t *is a function that assigns a temporal marker to each ternary edge.*

Therefore a meta graph of a resource provides additional contextual information regarding an entity at a given point in time. In this work we make use of the following features extracted from a resource meta graph:

- **Resource feature** (Res)**:** Consisting of the resource for which the semantic meta graph is derived. For example for the `dbp:Barack_Obama` resource.
- **Class Type features** (Cls)**:** Consisting of all classes appearing in the semantic meta graph of a resource that we derive from DBpedia. For example for the `dbp:Barack_Obama` resource these features include `dbo:OfficeHolder`.
- **Category features** (Cat)**:** Consisting of all resources of type `skos:Concept` appearing in the DBpedia semantic meta graph of an entity. For example for the `dbp:Barack_Obama` resource these features include `cat:Obama_family`.

[7] `http://freebase.com`
[8] `http://www.mpi-inf.mpg.de/yago-naga/yago/`

- **Property features (**$Prop$**):** Consisting of all properties appearing on the DBpedia-derived semantic meta graph of an entity. For example for the `dbp:Barack_Obama` resource these features include `foaf:givenName` and `dbo:writer`.

Therefore a document can be represented by the semantic features derived from the entities it contains. One approach to weight the semantic feature vector of a document is to use a frequentist approach, like the Semantic Feature Frequency (SFF)[3] weighting strategy which computes the frequency of a feature on a document applying a Laplace smoothing. This SFF will be our baseline for comparing the set of weighting strategies introduced in the following subsection.

4.3 Topic-Relevance Strategy for weighting features with DBpedia

Rather than characterising the relevance of a feature on a resource's graph (as in [3] [26]), here we aim to characterise the global relevance of a semantic feature to a given topic in DBpedia at a given point in time. For this we propose a novel set of semantic feature weighting strategies which rely on the semantic representation of a topic derived from DBpedia. As discussed in Section 3, a topic such as `War` can be represented by the collection of resources belonging to the `cat:War` category, and resources from its subcategories. This collection of resources build a topic-based graph structure that characterises this topic and evolves as new resources are added to the DBpedia graph.

The following strategies make use of a time-stamped DBpedia Topic graph to derive a feature's relative importance to this topic at a given time. When analysing the children to parent category relations we set the number of traversing steps to 2. In order to capture the relative importance of a feature to a given topic, we propose the following weighting strategies:

- ***Class-based Topic Relevance*** (Cls_W): Weights a type-feature f as the ratio of the number of distinct resources whose `rdf:type` is f and are labeled with categories appearing on the Topic graph, and the number of resources of `rdf:type` f derived from a DBpedia graph at time t (DB_t). For example to weight the type `dbo:OfficeHolder`[9] in the context of the Topic `War` we compute this weight as depicted in Figure 3.

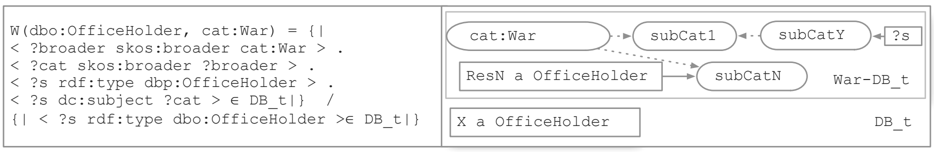

Fig. 3. Class Feature Weighting Strategy (Cls_W).

where DB_t[10] represents the DBpedia graph at time t. A higher weight means that the type feature f appears more often on resources derived from `cat:War`, therefore is more relevant to this Topic.

- ***Property-based Topic Relevance*** ($Prop_W$): Weights a property-feature f as the ratio of the number of distinct resources whose property is f and are labeled with categories appearing on the Topic graph; and the number of resources of type f derived from a DBpedia graph at time t (DB_t). For example to weigh the property

[9] dbo, qname for `http://dbpedia.org/ontology/`

[10] DBpedia graph snapshots are based on different DBpedia dumps described in section 5

dbProp:currency[11] in the context of the Topic War we compute this weight as depicted in Figure 4.

```
W(dbProp:currency, cat:War) =  {|
< ?broader skos:broader cat:War >.
< ?category skos:broader ?broader >.
< ?s dbProp:currency ?val >.
< ?s dc:subject ?category >∊ DB t|} /
{| < ?s dbProp:currency ?value >∊ DB t|}
```

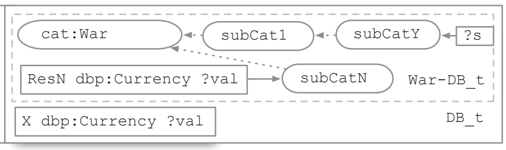

Fig. 4. Property Feature Weighting Strategy ($Prop_W$).

- ***Category-based Topic Relevance*** *(Cat_W)*: Weighs a category-feature f based on the number of resources appearing on sibling categories, which are also descendants of the main Topic category; divided by the number of resources belonging to the category and subcategories of the main Topic category derived from a DBpedia graph at time t (DB_t). For example to weight the type cat:Conflict in the context of the Topic War we compute this weight as described in Figure 5:

```
W(cat:Conflict, cat:War) = {|
< cat:Conflict skos:broader ?parent >.
< ?group skos:broader ?parent > .
< ?s dc:subject ?group > .
< ?broader skos:broader cat:War > .
< ?category skos:broader ?broader > .
< ?s dc:subject ?category > ∊ DB t|}/
{| < ?broader skos:broader cat:War >.
< ?category skos:broader ?broader > .
< ?s dc:subject ?category >∊ DB t|}
```

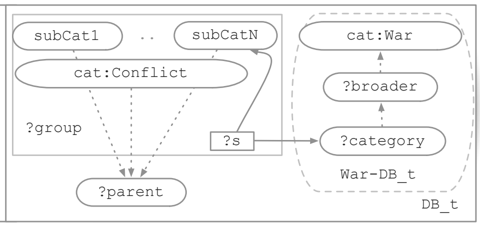

Fig. 5. Category Feature Weighting Strategy (Cat_W).

- ***Resource Relevance*** (Res_W): This weighting strategy does not make use of the topic graph, but rather characterises the relevance of a resource by comparing it to other resources. It is defined as the ratio of the number of resources which share this resource's categories and the number of resources in DBpedia labelled by a category derived from a DBpedia graph at time t (DB_t). For example to weight the resource dbp:Barack_Obama[12] we compute this weight as described in Figure 6:

```
W(dbp:Barack Obama) = {|
< ?category skos:broader ?broader > .
< ?s dc:subject ?category >
< dbp:Barack_Obama dc:subject ?category >
∊ DB t|} /
{| < ?s dc:subject ?cat >∊ DB t|}
```

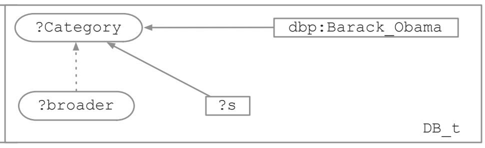

Fig. 6. Resource Feature Weighting Strategy (Res_W).

Once the semantic feature space of a corpus has been weighted based on the above weighting strategies, we integrate these weights into the feature representation of a tweet post by multiplying the number of times the feature appears on the document by the feature weight derived from the DBpedia graph ($\mathcal{DB}\_t$). Therefore the semantic feature f in a document x is weighted based on the frequency of a semantic feature f

[11] dbProp, qname for http://dbpedia.org/property/

[12] dbp, qname for http://dbpedia.org/resource/

in a document x with Laplace smoothing and the topic-relevance of the feature in the $\mathcal{DB}\_t$ graph:

$$W_x(f)_{\mathcal{DB}\_t} = [\frac{[N_x(f)_{\mathcal{DB}\_t} + 1}{|F| + \sum_{f' \in F} N_x(f')_{\mathcal{DB}\_t}}] * (W_{\mathcal{DB}\_t}(f))^{1/2} \quad (1)$$

where $N_x(f)$ is the number of times feature f appears in all the semantic meta-graphs associated with document x derived from the $\mathcal{DB}\_t$ graph ; F is the semantic features' vocabulary of the semantic feature type and $W_{\mathcal{DB}\_t}(f)$ is the weighting function corresponding to the semantic feature type computed based on the $\mathcal{DB}\_t$ graph.[13] This weighting function captures the relative importance of a document's semantic features against the rest of the corpus and incorporates the topic-relative importance of these features in the $\mathcal{DB}\_t$ graph.

4.4 Construction of Time-Dependent Topic Classifiers

To characterise the time-dependent impact on the decay in performance of a topic classifier we focus on the binary topic classification task in cross-epoch-based scenarios. In these scenarios the classifier that we train on a corpus from epoch $t-1$, is tested on a corpus on epoch t. We use our semantic graphs to characterise the two corpora, to verify our hypothesis that, as opposed to lexical features which are situation-dependent and can change progressively in time, semantic structures – including ontological classes and properties – can provide a more stable representation of a Topic in cross-epoch settings.

Following the weighting strategies in the previous section, the semantic feature representations of the $t-1$ corpus and the t corpus, are both generated from the DBpedia graph available at $t-1$. For example when applying a classifier trained on data from 2010, the feature space of a target test set from 2011 is computed based on the DBpedia version used for training the 2010-based classifier. This is in order to simulate the availability of resources in a DBpedia graph at a given time.[14]

5 Experimental Setup

In this section we introduce our datasets and present the experimental setting for evaluating the effectiveness of the proposed weighting strategies on a cross-epoch transfer learning task.

5.1 Dataset Description

Our datasets comprise two main collections: DBpedia and Twitter datasets. The DBpedia collection is comprised of four DBpedia dumps (3.6 to 3.9).[15] These dumps were installed on a Virtuoso server using separate named-graphs for each dump to facilitates dump-specific SPARQL queries. The DBpedia dumps allow us to extract semantic features for resources contained on a tweet, based on a specific DBpedia graph available at a particular epoch.

The Twitter datasets consist of a collection of Violence-related topics: Disaster_Accident, Law_Crime and War_Conflict. Each of these datasets comprises three epoch-based collections of tweets, corresponding to 2010, 2011, and 2013. The 2010 collection was gathered during November 2010 and December 2010 comprising over 1 million tweets. The 2011 collection was gathered during August 2011 also comprising over 1 million

[13] Notice that the square root on the proposed weight aids to emphasize this value, since the order of magnitude of this weight tends to be low.

[14] The comparison based on progressive availability of resources is future work

[15] General statistics of these dumps are available at http://wiki.dbpedia.org/Downloads39

tweets. Finally the 2013 collection was sampled during September of 2013 also comprising of over 1 million tweets. To generate our gold standard we first labelled these tweets using the topic labelling service from OpenCalais[16] which classifies a tweet into 18 different categories.[17] Then for each year we retrieved those tweets with labels corresponding to "Disaster & Accident", "Law & Crime" and "War & Conflict".

Based on a random selection of 10,000 tweets for each year of each Topic we used the AlchemyAPI service to extract entities. Then we performed a manual annotation based only on those tweets which contained at least one resource. We stop the manual annotation of a randomly sorted sample for each Topic for each year when reaching 1,000 tweets per topic per year, giving us a total of 9,000 tweets. In order to generate a negative set for each year, we used a 10,000 sample of the OpenCalais annotated set with tweets annotated with categories other than these three. We also pre-filtered tweets which contained at least one entity. Since in this work our aim is topic characterisation rather than violence detection we decided to keep balanced sets. Therefore for each year we kept a manual annotation of 1,000 tweets which are not related to any of these three topics. Based on the manual re-annotation of two annotators (computer science researchers) we achieved an averaged inter-annotator Kappa score of 73.5%. The final Twitter dataset therefore contained 12,000 annotated tweets.

In order to derive the lexical features, these datasets were preprocessed by first removing punctuation, numbers, non-alphabet characters, stop words, and links. We then performed Porter stemming [20] in order to reduce the vocabulary size. To generate the semantic features we used the disambiguated DBpedia links provided by AlchemyAPI. However since Alchemy is based on the most recent DBpedia dump, we resolved each disambiguated DBpedia resource to the DBpedia dump available at the time in which the tweet was created. Therefore for each document we only kept those entities which existed on the DBpedia dump available at the time in which the tweet was created.

The general statistics of these datasets including semantic features is summarised in Table 1. In this work we follow a frequency-based weighting strategy, which is a common approach in Information Retrieval. However here we report that only 26% of the lexical features in our Twitter dataset have frequency greater than 1 on a document. For the semantic feature spaces we have the following distributions: Cat-11%,Prop-94.7%,Res-1%,Cls-29% [18]. Notice that for each cross-time setting scenario presented in Section 6 where a classifier at time t is tested on a dataset at $t+1$, we recalculated the semantic features of the $t+1$ dataset to point back to the DBpedia graph available at time t.

Table 2, presents the top three lexical and semantic features ranked based on the SFF baseline strategy and based on our weighting strategies for the 2010 Law_Crime topic. The left column present the top semantic features ranked using our baseline (SFF) while the right column presents top features ranked using our semantic weighting strategies (SFG). Notice that while the frequency based strategy (SFF) seem to provide a representation specific to the current-situation modelling the Topic; the proposed SFG seem to provide a broader representation of the Topic based on the information derived from the DBpedia graph.

[16] OpenCalais, http://www.opencalais.com

[17] Full list of OpenCalais categories, http://www.opencalais.com/documentation/calais-web-service-api/api-metadata/document-categorization

[18] Averaged for the three topics and three years

		Unigram	Category	Properties	Resource	Class	tweets
D & A	2010	1,361	1,224	1,862	218	60	1,000
	2011	1,118	711	1,533	111	66	1,000
	2013	1,380	1,615	2,260	220	63	1,000
L & C	2010	1,427	1,577	1,795	213	65	1,000
	2011	1,012	870	1,698	111	70	1,000
	2013	1,288	1,530	2,202	208	104	1,000
W & C	2010	1,300	1,196	1,440	182	46	1,000
	2011	1,038	601	1,245	95	58	1,000
	2013	1,263	1,515	2,105	202	98	1,000
Neg	2010	1,634	2,044	2,167	229	86	1,000
	2011	1,244	1,562	2,080	160	101	1,000
	2013	1,194	1,896	2,048	162	114	1,000

Table 1. Statistics of the lexical and semantic features extracted for the Disaster_Accident (D & A), Law_Crime (L & C), War_Conflict (W & C), and Negative (Neg) tweet collections. The reported statistics for Unigrams is after preprocessing.

		2010-SFF	2010-SW
Law_Crime	Lex	wikileak, arrest, law	wikileak, arrest, law
	Cat	cat:Living_People, cat:Liberal_democracies, cat:G20_nations	cat:Living_People, cat:Theft, cat:Commercial_crimes
	$Prop$	dc:subject, foaf:name, dbpProp:leaderName	foaf:page, rdf:label, dbpProp:name
	Res	dbpr:United_States, dbpr:Julian_Assange, dbpr:Wikileaks	dbpr:Marc_Emery, dbpr:Erik_Bornmann, dbpr:Reggie_Bush
	Cls	dbp:Place, gml:_Feature, dbp:PopulatedPlace	dbp:Work, dbp:Criminal, dbp:Person

Table 2. An extract of the feature space of the Law_Crime Topic of 2010. We also present the top three features for unigram. The qualified names used in this table are mapped as follows: [dbp, http://dbpedia.org/ontology], [dbc, http://dbpedia.org/resource/Category/], [dbpr, http://dbpedia.org/resource/], [dbpProp, http://dbpedia.org/property/], [gml, http://www.opengis.net/gml/], [skos, http://www.w3.org/2004/02/skos/core#], [foaf, http://xmlns.com/foaf/0.1/], [dc, http://purl.org/dc/terms/subject].

5.2 Experimental Setting

To assess the features temporal impact on a classification task we use as a baseline the performance of a topic classifier trained and tested on an epoch t. In this case we assess performance differences when a classifier is tested on future epochs as we described in Section 4.4. We use the standard weighting strategies as a baseline (i.e., TF-IDF for BoW and SFF for semantic features[19]) to compare against the weighting ones introduced in section 4.2.

To test whether semantic features can aid on this cross-epoch transfer learning task, we performed the following series of experiments. For each topic we built supervised topic classifiers using the independent feature types (i.e., bag of words features [BoW], semantic features –class [Cls], property [Prop], category [Cat], resource [Res]–) and the merged features (i.e., joint-semantic features, [Sem], and the BoW + semantic features [All]). In this collection of classifiers features were weighted based on our baseline weighting strageties: TF-IDF for the BoW features and SFF for the semantic features

[19] We used the SSF weighting strategy in order to have a one to one comparison based on semantic feature types. This is the reason why we did not include the class-property co-occurrence frequency [3] strategy in our baseline

($\_SFF$). We also generated the same set of classifiers but this time using the SFG weighting strategies ($\_SFG$) introduced in Section 4. We also generated merged settings, here Sem_{SFF} and Sem_{SFG} correspond to classifiers trained on joint semantic features weighted with SFF and SFG respectively. Sem_{joint} refers to classifiers using all semantic features weighted with the $SFF+SFG$ setting. The All classifiers are based on the semantic + BoW settings; the subscript indicates the weighting scheme.

6 Experimental Results

In this section we address the following questions: *Do semantic features built from DBpedia Graphs aid on a cross-epoch transfer learning task for the topic classification of Tweets? if so, to what extent can these semantic features help the classification task?* In our experiments we used Support Vector Machine (SVM) [6] with polynomial kernel classifiers. All the experiments reported here were conducted using a 10-fold cross validation setting [22][14].

6.1 Evaluation of Semantic Features on Same-Epoch Scenarios

In order to assess the benefit of using semantic features in topic classification, we start by studying their role when a topic classifier is trained and tested on the same epoch.

Table 3 shows the results of topic classifiers trained and tested on the same years and datasets, using (1) BoW features (i.e., lexical features); (2) baseline semantic features weighted based on SFF (Section 4.2); (3) semantic features with our graph-based weighting strategies (SFG, Section 4.3); (4) using joint semantic features (Sem); and (5) using the joint BoW and semantic features (All).

Results show that in same-epoch scenarios, BoW features outperform all semantic features in topic classification. They also show that while results with SFF are better than with SFG in almost all cases, their joint use outperform the SFF baseline in P. These are interesting, but unsurprising results. This is because the training and classification are done on the same dataset and epoch, and hence the current data content should be more representative of the topic. However this set of same-year results become our baseline against cross-epoch settings (where these classifiers are tested on future epochs).

6.2 Evaluation of Semantic Features on Cross-Epoch Scenarios

Now we study the performance of our BoW and semantic features when the training is done on one epoch and the classification is applied to another. This will help us understand how these features decay across epochs.

Table 4 presents results for three cross-epoch scenarios for the Disaster_Accident (Dis_Acc) topic. Each X-Y column refers to the performance of a classifier trained on epoch X and tested on epoch Y. The last column presents the average results of this topic across these cross-epoch scenarios. Comparing the performance of the Dis_Acc 2010 classifier (Table 3) with the 2010-2011, and 2010-2013 results (Table4) we observe a consistent drop in F measure when this Dis_Acc 2010 TC is applied using BoW features. The same occurs when comparing performance of Dis_Acc 2011 when applied to 2013. Moreover we observe that for this Topic all individual semantic features types -weighted with SFF, SFG and SFF+SFG (Joint)–, consistently outperform the BoW baseline in F-measure. When analysing the overall contribution of semantic features we observe that in average all semantic features weighted with SFG (Sem_{SFG}) significantly improve P when compared to the SFF baseline (Sem_{SFF}) (t-test with $\alpha < 0.01$), while consistently improve F-measure when compared to the averaged BoW features (t-test with $\alpha < 0.01$).

		Dissaster_Acc			Law_Crime			War_Conflict		
		P	R	$F1$	P	R	$F1$	P	R	$F1$
	BoW	0.855	0.809	**0.831**	0.776	0.756	**0.765**	0.868	0.821	**0.844**
2010	Cat_{SFF}	0.740	0.661	0.697	0.639	0.663	0.650	0.781	0.712	0.744
	Cat_{SFG}	0.744	0.546	0.629	0.731	0.458	0.562	0.797	0.698	0.743
	Cat_{joint}	0.769	0.608	0.678	0.716	0.508	0.594	0.793	0.686	0.735
	$Prop_{SFF}$	0.720	0.646	0.680	0.612	0.671	0.639	0.749	0.694	0.720
	$Prop_{SFG}$	0.711	0.618	0.659	0.584	0.697	0.635	0.735	0.678	0.705
	$Prop_{joint}$	0.734	0.623	0.673	0.588	0.685	0.632	0.759	0.679	0.716
	Res_{SFF}	0.773	0.627	0.692	0.724	0.569	0.637	0.812	0.720	0.762
	Res_{SFG}	0.776	0.567	0.654	0.749	0.499	0.599	0.812	0.656	0.725
	Res_{joint}	0.775	0.600	0.675	0.751	0.510	0.607	0.821	0.677	0.741
	Cls_{SFF}	0.637	0.631	0.633	0.552	0.629	0.583	0.688	0.595	0.637
	Cls_{SFG}	0.632	0.608	0.619	0.582	0.486	0.527	0.666	0.573	0.614
	Cls_{joint}	0.635	0.606	0.619	0.583	0.510	0.542	0.684	0.584	0.628
	Sem_{SFF}	0.746	0.700	0.720	0.639	0.683	0.659	0.782	0.740	0.760
	Sem_{SFG}	0.685	0.773	0.725	0.629	0.738	0.678	0.757	0.740	0.748
	Sem_{joint}	0.777⋆	0.652	0.708	0.716⋆	0.553	0.623	0.795⋆	0.715	0.752
	All_{SFF}	0.817	0.791	0.803	0.761	0.766	0.763	0.851	0.830	0.840
	All_{SFG}	0.807	0.814	0.809	0.764	0.789	0.776	0.847	0.837	0.841
	All_{joint}	0.829	0.769	0.797	0.782	0.726	0.752	0.860	0.814	0.836
	BoW	0.899	0.853	**0.875**	0.868	0.808	**0.836**	0.905	0.860	**0.882**
2011	Cat_{SFF}	0.841	0.735	0.784	0.830	0.697	0.756	0.881	0.817	0.847
	Cat_{SFG}	0.848	0.698	0.765	0.849	0.681	0.755	0.881	0.798	0.837
	Cat_{joint}	0.852	0.724	0.782	0.842	0.683	0.753	0.879	0.806	0.840
	$Prop_{SFF}$	0.815	0.722	0.765	0.763	0.661	0.706	0.856	0.806	0.830
	$Prop_{SFG}$	0.812	0.714	0.759	0.780	0.652	0.709	0.872	0.797	0.832
	$Prop_{joint}$	0.825	0.716	0.766	0.778	0.656	0.711	0.856	0.797	0.824
	Res_{SFF}	0.856	0.736	0.791	0.849	0.706	0.770	0.886	0.810	0.846
	Res_{SFG}	0.882	0.702	0.781	0.871	0.655	0.746	0.896	0.779	0.832
	Res_{joint}	0.880	0.699	0.779	0.865	0.679	0.760	0.893	0.788	0.837
	Cls_{SFF}	0.714	0.712	0.712	0.700	0.616	0.653	0.824	0.773	0.797
	Cls_{SFG}	0.716	0.710	0.712	0.705	0.584	0.636	0.814	0.761	0.786
	Cls_{joint}	0.714	0.709	0.711	0.697	0.613	0.650	0.811	0.761	0.784
	Sem_{SFF}	0.814	0.761	0.786	0.805	0.729	0.764	0.861	0.824	0.841
	Sem_{SFG}	0.807	0.767	0.786	0.774	0.727	0.748	0.855	0.823	0.838
	Sem_{joint}	0.831⋆	0.744	0.784	0.824⋆	0.714	0.764	0.871⋆	0.809	0.838
	All_{SFF}	0.876	0.846	0.861	0.843	0.804	0.822	0.882	0.844	0.862
	All_{SFG}	0.884	0.858	0.870	0.846	0.814	0.829	0.884	0.853	0.868
	All_{joint}	0.878	0.844	0.860	0.856	0.787	0.819	0.887	0.836	0.860
	BoW	0.862	0.806	**0.833**	0.875	0.832	**0.852**	0.870	0.808	**0.838**
2013	Cat_{SFF}	0.774	0.687	0.727	0.798	0.682	0.734	0.756	0.657	0.701
	Cat_{SFG}	0.807	0.625	0.704	0.817	0.634	0.713	0.780	0.606	0.681
	Cat_{joint}	0.791	0.658	0.717	0.826	0.644	0.723	0.788	0.622	0.694
	$Prop_{SFF}$	0.762	0.680	0.718	0.771	0.682	0.723	0.742	0.657	0.696
	$Prop_{SFG}$	0.748	0.657	0.699	0.772	0.680	0.722	0.753	0.665	0.705
	$Prop_{joint}$	0.768	0.672	0.716	0.777	0.673	0.720	0.765	0.665	0.711
	Res_{SFF}	0.788	0.660	0.718	0.821	0.663	0.733	0.787	0.611	0.687
	Res_{SFG}	0.800	0.623	0.700	0.836	0.634	0.720	0.804	0.606	0.690
	Res_{joint}	0.806	0.614	0.696	0.836	0.632	0.719	0.813	0.578	0.673
	Cls_{SFF}	0.707	0.659	0.680	0.745	0.657	0.697	0.694	0.653	0.671
	Cls_{SFG}	0.717	0.609	0.657	0.750	0.647	0.693	0.704	0.649	0.672
	Cls_{joint}	0.716	0.634	0.671	0.748	0.658	0.699	0.702	0.674	0.686
	Sem_{SFF}	0.767	0.719	0.741	0.772	0.725	0.747	0.751	0.706	0.728
	Sem_{SFG}	0.741	0.762	0.751	0.754	0.755	0.754	0.736	0.743	0.739
	Sem_{joint}	0.778⋆	0.694	0.733	0.803⋆	0.681	0.736	0.770⋆	0.656	0.708
	All_{SFF}	0.832	0.799	0.814	0.844	0.800	0.821	0.836	0.804	0.819
	All_{SFG}	0.837	0.824	0.830	0.845	0.821	0.832	0.835	0.819	0.827
	All_{joint}	0.844	0.781	0.811	0.854	0.779	0.814	0.840	0.764	0.799

Table 3. Performance of the classifiers trained and tested on the same epoch. The classifiers where applied on testsets weighted based on the classifier weighting scheme. The values highlighted in bold correspond to the best results obtained in F measure for each topic and each year. A ⋆ denotes that the P-measure of a given weighted feature significantly outperforms the corresponding SFF baseline. Significance levels: p-value < 0.01

		2010-2011			2010-2013			2011-2013			Average		
		P	R	$F1$	P	R	$F1$	P	R	$F1$	P	R	$F1$
	BoW	0.807	0.526	0.634	0.773	0.350	0.481	0.857	0.155	0.261	0.812	0.343	0.458
Disaster_Acc	Cat_{SFF}	0.721	0.650	0.683	0.696	0.443	0.539	0.808	0.389	0.524	0.741	0.494	0.582
	Cat_{SFG}	0.766	0.613	0.677	0.766	0.483	0.592	0.809	0.468	0.592	0.780	0.521	0.620
	Cat_{Joint}	0.798	0.645	0.713	0.734	0.310	0.434	0.818	0.381	0.518	0.783	0.445	0.555
	$Prop_{SFF}$	0.708	0.631	0.665	0.656	0.486	0.557	0.718	0.387	0.502	0.694	0.501	0.574
	$Prop_{SFG}$	0.689	0.676	0.681	0.668	0.489	0.564	0.750	0.453	0.564	0.702	0.539	0.603
	$Prop_{Joint}$	0.724	0.652	0.686	0.686	0.480	0.564	0.717	0.352	0.470	0.709	0.494	0.573
	Res_{SFF}	0.794	0.756	0.774	0.723	0.438	0.544	0.770	0.317	0.445	0.762	0.503	0.587
	Res_{SFG}	0.818	0.752	0.783	0.791	0.486	0.599	0.786	0.299	0.423	0.798	0.512	0.601
	Res_{Joint}	0.806	0.754	0.779	0.765	0.477	0.586	0.788	0.284	0.409	0.786	0.505	0.591
	Cls_{SFF}	0.684	0.701	0.691	0.666	0.667	0.665	0.705	0.638	0.669	0.685	0.668	0.675
	Cls_{SFG}	0.679	0.700	0.689	0.663	0.657	0.660	0.700	0.644	0.670	0.680	0.667	0.673
	Cls_{Joint}	0.688	0.704	0.695	0.668	0.656	0.661	0.699	0.640	0.667	0.685	0.666	0.674
	Sem_{SFF}	0.720	0.683	0.700	0.699	0.493	0.578	0.814	0.411	0.545	0.744	0.529	0.607
	Sem_{SFG}	0.755⋆	0.599	0.668†	0.776⋆	0.371	0.501†	0.816⋆	0.333	0.472†	0.782⋆	0.434	0.547†
	Sem_{Joint}	0.781⋆	0.623	0.693†	0.720⋆	0.402	0.515†	0.815†	0.313	0.451†	0.772⋆	0.446	0.553†
	All_{SFF}	0.768	0.555	0.642	0.771	0.428	0.549	0.845	0.205	0.330	0.565	0.396	0.507
	All_{SFG}	0.791⋆	0.546	0.644†	0.724⋆	0.388	0.505†	0.850⋆	0.210	0.335†	0.788⋆	0.381	0.494†
	All_{Joint}	0.798⋆	0.527	0.632	0.791⋆	0.372	0.504†	0.844⋆	0.168	0.279†	0.811⋆	0.355	0.471†

Table 4. Presents results for the cross-epoch scenarios for the Disaster_Accident topic. A ⋆ denotes that the P-measure of the shaded cell significantly outperforms their corresponding SFF baseline. A † denotes that the F-measure of a weighted feature outperforms the BoW baseline. Significance levels: p-value < 0.01.

To compare the benefit of the proposed weighting strategies across all topics we computed the averaged P, R, F1 across epochs for each Topic. These averages, presented in Table 5, show that in average the merged SFG (Sem_{SFG} features significantly outperforms the merged SFF (Sem_{SFF}) features (t-test with $\alpha < 0.01$) classifiers by 4.3%. These results also show that on cross-epoch scenarios, on average, some individual semantic features-based classifiers outperform the BoW classifier in F-measure obtaining a maximum increment of 16.37% (t-test with $\alpha < 0.01$) when using the Cls_{joint} weighted feature. Moreover Class semantic features(Cls) alone (ClsSFF, ClsSFG, ClsJoint) in average consistently outperform BoW in F with a gain of 12.5% for all cross-epoch scenarios for all three topics. This demostrates that the use of Cls semantic features alone compared to lexical features is benefitial in characterising a topic in time.

We also observe that when incorporating BoW to the semantic feature space – extended feature representation of a document, where a tweet is represented using its lexical+semantic features – we consistently outperform the BoW baseline for the three joint settings (All_{SFG}, All_{SFG}, All_{joint}) with the highest F-measure achieved by the All_{SFG} classifier. This setting significantly outperforms the BoW classifier in F-measure by 1.6% (t-test with $\alpha < 0.01$) while providing the best precision across semantic features. This positive increment indicates that the incorporation of external knowledge (DBpedia-graph) in the cross-epoch transfer learning task is beneficial when applied jointly with document derived weighting strategies.

We analyse the relevance decay of features based on performance gain on the cross-epoch scenarios. These is calculated by comparing the cross-scenario performance of each classifier against the performance of the corresponding classifier on the same-year scenario (e.g. 2010-2011 compared against 2010). The heatmap to the left in Figure 7 presents our results for all features. The heatmap to the right in Figure 7 presents

	Disaster_Acc			Law_Crime			War_Conflict			Average		
	P	R	$F1$	P	R	$F1$	P	R	$F1$	P	R	$F1$
BoW	0.812	0.343	0.458	0.739	0.549	0.620	0.873	0.394	0.531	0.808	0.429	0.536
Cat_{SFF}	0.741	0.494	0.582	0.641	0.479	0.537	0.774	0.325	0.453	0.719	0.433	0.524
Cat_{SFG}	0.780	0.521	0.620	0.769	0.432	0.549	0.803	0.350	0.480	0.784	0.434	0.55
Cat_{Joint}	0.783	0.445	0.555	0.766	0.426	0.542	0.777	0.280	0.406	0.775	0.383	0.501
$Prop_{SFF}$	0.694	0.501	0.574	0.604	0.445	0.504	0.755	0.411	0.506	0.684	0.452	0.528
$Prop_{SFG}$	0.702	0.539	0.603	0.596	0.468	0.509	0.731	0.391	0.460	0.676	0.460	0.524
$Prop_{Joint}$	0.709	0.494	0.573	0.618	0.462	0.518	0.767	0.383	0.487	0.698	0.446	0.526
Res_{SFF}	0.762	0.503	0.587	0.756	0.473	0.578	0.773	0.338	0.466	0.764	0.438	0.544
Res_{SFG}	0.798	0.512	0.601	0.757	0.428	0.539	0.771	0.337	0.448	0.775	0.426	0.529
Res_{Joint}	0.786	0.505	0.591	0.761	0.413	0.528	0.786	0.307	0.432	0.777	0.408	0.517
Cls_{SFF}	0.685	0.668	0.675	0.626	0.679	0.647	0.764	0.599	0.660	0.692	0.649	0.660
Cls_{SFG}	0.680	0.667	0.673	0.668	0.617	0.640	0.724	0.632	0.661	0.691	0.638	0.658
Cls_{Joint}	0.685	0.666	0.674	0.669	0.645	0.656	0.761	0.608	0.664	0.705	0.640	0.665
Sem_{SFF}	0.744	0.529	0.607	0.603	0.457	0.509	0.778	0.329	0.459	0.708	0.438	0.525
Sem_{SFG}	0.782	0.434	0.547	0.710	0.384	0.494	0.762	0.302	0.431	0.751	0.373	0.490
Sem_{Joint}	0.772	0.446	0.553	0.734	0.397	0.512	0.743	0.369	0.490	0.75	0.404	0.518
All_{SFF}	0.565	0.396	0.507	0.709	0.507	0.586	0.819	0.387	0.520	0.774	0.43	0.537
All_{SFG}	0.788⋆	0.381	0.494	0.756⋆	0.523	0.613	0.859⋆	0.411	0.550	0.801⋆	0.438	0.552†
All_{Joint}	0.811	0.355	0.471	0.762	0.471	0.578	0.795	0.449	0.571	0.789⋆	0.425	0.540†

Table 5. Average results for the cross-epoch scenarios for each topic. The last column present the average results of all three topics. A ⋆ denotes that the P-measure of the shaded cell significantly outperforms their corresponding SFF baseline. A † denotes that the F-measure of a weighted feature outperforms the BoW baseline. Significance levels: p-value < 0.01.

the averaged gain on BoW for three cross-epochs for each Topic. The heatmap to the left presents average gain on F-measure on a cross-scenario compared against its corresponding same-year scenario classifier. A higher value indicates that the feature adapts better (i.e. lower decay) in a cross-epoch setting, while a lower value indicates that on average the feature is less relevant for a topic on a cross-epoch setting. The heatmap to the right presents the average gain on BoW F-measure on a cross-scenario compared against its corresponding BoW gain on a same-year scenario classifier. Here a higher value indicates that a feature adapts better than the BoW on a cross-epoch setting, while a lower value indicate otherwise. Here we observe that on a cross-epoch setting the Cls semantic features are highly relevant for the cross-epoch learning task. Moreover based on these results, these semantic feature appears to provide more stable (i.e. lower decay) information than the one provided by the BoW. In this case the Cls_{joint} exhibits a gain which exceeds on over 7% the BoW one.

Finally to analyse the gain over BoW on the cross-epochs, we computed gain differences on the BoW F-measure obtained by each feature, and compared it with the one of the same-year scenarios. These results indicate that on average the Cls features exhibit a lower decay when compared to the BoW providing a more stable F-measure on the cross-epoch scenarios.

7 Discussion

In this paper we introduce a novel approach to the cross-epoch transfer learning task. This approach proposes the use of semantic features as a more stable representation of a topic over time. While the proposed set of weighting strategies is based on heuristics, other weighting strategies could be studied in future work. Such strategies could be enhanced with methods and results from work on ontology and linked data searching, ranking, and summarisation. Also other lexical features (e.g. part-of-speech) and

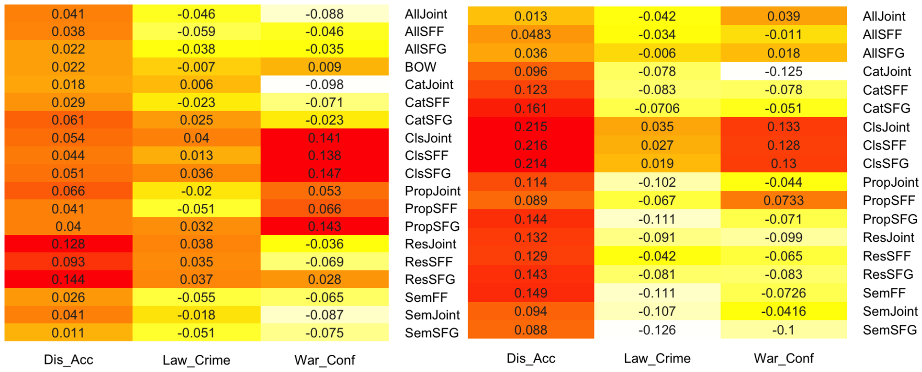

Fig. 7. Averaged gain on BoW for three cross-epochs for each Topic

structure information (e.g. WordNet)[20] could be used along with semantic features to improve performance.

The limited availability of annotated datasets spanning across longer periods of time made us focus only on a range of three different epochs. This work could be further expanded by considering longer periods of time, and by experimenting with different type of topics. So far we have demostrated that for the violence-related topics the Cls feature exhibited the lowest relevance decay on the transfer learning task. For these topics some individual features were less performing that others. However further research is necessary to understand what makes a semantic feature a good option for the cross-epoch modeling task depending on the type of topic.

8 Conclusions and Future Work

In this paper we proposed the use of semantic features to approach the cross-epoch transfer learning task for topic classification of tweets. Moreover we introduced a framework which proposes to enrich semantic features by incorporating information derived from an external knowledge source. The framework introduced a set of weighting strategies which calculates the relevance of features from time-stamped topic graphs extracted from DBpedia. Our results showed that semantic features are much slower to decay than other features, and that they can improve performance upon traditional BoW-based classifiers in cross-epoch scenarios. Furthermore, results showed that the proposed strategies improve performance upon our baseline while outperforming F-measure upon BoW features. These results demonstrate the feasibility of the use of semantic features in epoch-based transfer learning tasks. This opens new possibilities for the research of concept drift tracking for transfer learning based on existing Linked Data sources. Future work includes the comparison of semantic feature based transfer learning with other state of the art transfer learning approaches based on lexical features.

Acknowledgements

This work was partially supported by European Project Sense4us (611242).

[20] WordNet, http://wordnet.princeton.edu/

References

1. J. Blitzer, R. McDonald, and F. Pereira. Domain adaptation with structural correspondence learning. In *Proc. Conf. on EMNLP*, 2006.
2. A. E. Cano, Y. He, L. K., and Z. J. A weakly supervised bayesian model for violence detection in social media. In *Proc. 6th IJCNLP2013*, 2013.
3. A. E. Cano, A. Varga, M. Rowe, F. Ciravegna, and Y. He. Harnessing linked knowledge source for topic classification in social media. In *Proc. 24th ACM Conf. on Hypertext and Social Media*, Paris, France, 2013.
4. R. Caruana. Multitask learning. 28(1):41–75, 1997.
5. G. H. Chen, S. Nikolov, and D. Shah. A latent source model for nonparametric time series classification. In *Advances in Neural Information Processing Systems.*, 2013.
6. N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines: and Other Kernel-Based Learning Methods*. Cambridge University Press, 2000.
7. I. Daumé. Frustratingly easy domain adaptation. In *Proceedings of the 2007 ACL*, 2007.
8. Q. Diao, J. Jiang, F. Zhu, and E.-P. Lim. Finding bursty topics from microblogs. In *Proc. 50th Annual Meeting of the ACL*, Jeju Island, Korea, 2012.
9. A. Dries and U. Rückert. Adaptive concept drift detection. *Stat. Anal. Data Min.*, 2(56), 2009.
10. Y. Genc, Y. Sakamoto, and J. V. Nickerson. Discovering context: classifying tweets through a semantic transform based on wikipedia. In *Proc. of the 6th Int. Conf. on FAC*, 2011.
11. Y. He. Incorporating sentiment prior knowledge for weakly supervised sentiment analysis. *ACM Transactions on Asian Language Information Processing*, 11(2):4:1–4:19, June 2012.
12. Y. He, C. Lin, W. Gao, and K.-F. Wong. Tracking sentiment and topic dynamics from social media. In *Proc. of the Sixth Int. Conf. on Weblogs and Social Media, Dublin, Ireland*, 2012.
13. K. S. Jones, S. Walker, and S. E. Robertson. A probabilistic model of information retrieval: Development and comparative experiments. *Inf. Process. Manage.*, 36(6):779–808, 2000.
14. R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proc. 14th IJCAI - Volume 2*, 1995.
15. W. Lv, W. Xu, and J. Guo. Transfer learning in classification based on semantic analysis. In *2nd Int. Conf. on ICCSNT*, 2012.
16. N. Milikic, J. Jovanovic, and M. Stankovic. Discovering the dynamics of terms semantic relatedness through twitter. In *Proceedings, 1st Workshop on #MSM2011*, 2011.
17. O. Muñoz García, A. García-Silva, O. Corcho, M. de la Higuera Hernández, and C. Navarro. Identifying Topics in Social Media Posts using DBpedia. In *Proc. of the NEM Summit*, 2011.
18. S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Trans. on Knowl. and Data Eng.*, 22(10):1345–1359, Oct. 2010.
19. X.-H. Phan, L.-M. Nguyen, and S. Horiguchi. Learning to classify short and sparse text & web with hidden topics from large-scale data collections. In *Proc. 17th Int. Conf. on World Wide Web*, WWW '08, Beijing, China, 2008.
20. M. Porter. An algorithm for suffix stripping. *Program*, 14(3), 1980.
21. H. Saif, M. Fernández, Y. He, and H. Alani. Senticircles for contextual and conceptual semantic sentiment analysis of twitter. In *ESWC, Crete, Greece*, volume 8465, 2014.
22. S. L. Salzberg and U. Fayyad. On comparing classifiers: Pitfalls to avoid and a recommended approach. *Data Mining and Knowledge Discovery*, pages 317–328, 1997.
23. Y. Song, H. Wang, Z. Wang, H. Li, and W. Chen. Short text conceptualization using a probabilistic knowledgebase. In *Int. Joint. Conf. of AI (IJCAI)*. IJCAI/AAAI, 2011.
24. B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas. Short text classification in twitter to improve information filtering. In *Proc. of the Int. ACM SIGIR*, 2010.
25. S. Thrun. Is learning the n-th thing any easier than learning the first? In *Advances in Neural Information Processing Systems*, 1996.
26. A. Varga, A. Cano, M. Rowe, F. Ciravegna, and Y. He. Linked knowledge sources for topic classification of microposts: A semantic graph-based approach. *JWS: Science, Services and Agents on the WWW*, 2014.
27. X. Zhao, B. Shu, J. Jiang, Y. Song, H. Yan, and X. Li. Identifying event-related bursts via social media activities. In *Proc.of the Joint Conference on EMNLP*, Jeju Island, Korea, 2012.