OSLa Simões, Barreiro, Santos, Sousa-Silva & Tagnin (eds.) *Linguística, Informática e Tradução: Mundos que se Cruzam*, Oslo Studies in Language 7(1), 2015. 379–395. (ISSN 1890-9639 / ISBN 978-82-91398-12-9)
http://www.journals.uio.no/osla

THE IDENTIFICATION OF INDICATORS OF SENTIMENT USING A MULTI-VIEW SELF-TRAINING ALGORITHM

BRETT DRURY AND ALNEU DE ANDRADE LOPES

RESUMO

Este artigo apresenta um algoritmo de "multi-view self-training" , que identifica os indicadores de sentimento por: 1. extração relações causais, 2. As relações causais classificação em uma categoria sentimento, 3. agrupamento causas comuns e 4. atribuindo categorias sentimento a causas comuns para criar um distribuição sentimento para cada causa comum. Uma avaliação manual global da estratégia descobriu que ele tinha uma precisão de 70,00%.

[1] INTRODUCTION

Sentiment analysis has become an increasingly popular area of research. Sentiment analysis typically relies upon the detection of words that have a sentiment orientation. Sentiment analysis is used in time dependent tasks such as reputation management and stock trading. Reputation management identifies positive or negative in documents published on the Internet to gauge a value of a brand. Sentiment analysis in stock trading identifies positive, negative or neutral statements in news or blog posts to identify buy or sell signals for specific stocks or financial indexes. These tasks are time dependent because they rely upon sentiment to make inferences about future events. For example, profit warnings or sales figures. Once the event has happened, information related to the event is worthless. In time dependent sentiment analysis the further ahead in time sentiment about a future can be identified the more valuable the information.

This paper presents an algorithm for identifying indicators of sentiment. Indicators of sentiment for the purposes of this paper are noun phrases that indicate the existence of sentiment at sometime in the future.

The algorithm relies upon the detection of causal relations and the sentiment classification of the effect part of the causal relation. The algorithm groups together common causes and the associated sentiment classifications. The sentiment classifications are aggregated into a probability distribution. This sentiment probability distribution is an indicator of future sentiment implied by a mention of a cause in a text.

[2] RELATED WORK

The related work will discuss the following: causation in text, causal relation extraction, sentiment classification and prediction of future texts from information in past documents.

[2.1] *Causation in text*

Causal relations in text can be seen as relation that exists between two events if one event is the cause of the other (Altenberg 1984). Altenberg (1984) stated that three conditions must exist before a causative relation can exist in written or spoken language. The three conditions are: 1. encapsulate the two members of the relationship, 2. express the type of relationship between the relation's members and 3. identify the members in a coherent sequence. An alternate definition of causative relation was provided by Baron (1974) who stated: "Causation is a relationship between two states of affairs, X at time T_1 and X' at time T_2, and a cause Z that provides the necessary conditions for causing the change from X to X'". Baron (1974) provided four areas that should be considered when analyzing causative grammar: 1. what it is represented by the causative relation, 2. what mechanisms does the language have to represent causation, 3. what level in the grammar is the causation represented and 4. what syntactic/semantic parameters define the relationship between elements in causative constructions (Baron 1974). Baron (1974) further states that causation can be seen as a relation between entire propositions and/or sentences.

Two types of causation in text can be considered: explicit and implicit. Explicit causation is when the causative link is explicitly stated, for example in the generalization for causative verbs, $NP\,V\,NP$[1], that was provided by Levin (1993). An example of explicit causation that fits the $NP\ V\ NP$ pattern is "Smoking causes cancer.". Implicit causation is when the causal link is implied, for example, "The sun was bright and I was sweating". The implied cause the action of sweating is the warmth of the sun.

[2.2] *Causal relation extraction*

The causal relation extraction can be grouped into general methods: manual and automatic. Manual methods rely upon manually identified characteristics of language, typically patterns, to detect a causative relation. The automatic approaches tend to be supervised machine learning strategies. Supervised learning strategies are methods where labelled data is used to induce a classification model that is used to identify causal relations in unlabelled text.

[1] NP = Noun Phrase, V = Verb

Manual Approaches

A simple approach for manual strategies is to use hand crafted patterns. These patterns are typically created by human experts and can be domain specific, that can't be generalized to other domains. In addition the rule construction process can be a time consuming process. There were a number of approaches that relied upon domain knowledge and hand-crafted rules. One of the earliest examples found in the literature was by Kaplan (1991). His system had a pipeline that had several stages that were: 1. hand coded propositional representational parser, 2. semantic analysis component, 3. causal analysis and 4. knowledge base acquisition. Each stage is dependent upon the previous stage. The causal analysis component creates a causal chain of events based upon the output of the semantic analysis component (SAC). The output of the SAC are a series of concept frames that are represented as structured inheritance network. The root node of the network is known as "thing", and the sub-nodes can be members of one of the following classes: objects, actions, or relationships. The causal chain is constructed by using an event seed pair, for example, "air rising" and "air cooling". The effect part of the pair is used as a part of the next causal pair. This process continues until no more causal pairs can be made. The detection of causal pairs is achieved with "propositional clues". Joskowicz et al. (1989) identified causal links between messages generated by equipment installed in navy ships. This approach also relied upon a manual and domain specific approach.

Machine Learning

A popular supervised approach to extract causative relations is to use a sequence classification strategy. There are a number of machine learning methods that can be used in sequence classification strategies, for example Hidden Markov Models (HMM) and Maximum Entropy Markov Models (MEMM). The research literature indicates that one of the most common methods for causal relation extraction are Conditional Random Fields (CRF). Mehrabi et al. (2013) used CRFs in a supervised strategy to extract causative relations from texts about the Geriatric Care domain. The authors used the following features: tokens, token categories, prefix and suffixes, and Part Of Speech (POS) tag. The CRF had three possible labels: cause, effect and out.

Riaz & Girju (2014) used verbs and nouns as features for a classifier[2]. The features were grouped as: lexical, semantic and structural. Lexical features were described as "verb, lemma of verb, noun phrase, lemma of all words of noun phrase, head noun of noun phrase, lemmas of all words between verb and head noun of noun phrase.". The semantic features used were the nine noun hierarchies of WordNet. The structural features were the subject and object of a verb.

[2] The authors describe the classifier as a "basic supervised classifier".

[2.3] *Sentiment Analysis*

There are different types of sentiment analysis, for example: extraction of sentiment lexicons (fine grained) and classification (document level). This related work will concentrate on sentiment classification because it is directly related to the work described in this paper. Sentiment classification treats sentiment as a classification task that assigns a document to a category, typically: negative, neutral or positive. A common approach is to use machine learning (Pang et al. 2002). Machine learning uses training data to induce a classification model. The model is then used to classify unlabelled instances into the aforementioned categories. Labelled data for sentiment classification can be imbalanced with one category comprising the majority of the data-set (Drury & Lopes 2014). There are a number of strategies to reduce the effect of imbalanced data for sentiment classification, and balancing by oversampling seems to be the most effective for imbalanced Portuguese sentiment data (Drury & Lopes 2014).

Manually labelling data can be a time consuming task, consequently there has been a number of approaches that use semi-supervised learning.[3] Semi-supervised learning uses labelled and unlabelled data to produce a model from a classifier. One semi-supervised strategy for sentiment classification is self-training (He & Zhou 2011). Self-training induces a model from labelled instances and unlabelled data in an iterative way. In each iteration, high confidence classifications are added to the labelled data. At the end of an iteration a new model is induced from the new training data, and the process is continued. The process stops when there are no new instances added to the training data. Self-training can often produce worse results than supervised learning (Drury et al. 2011). This is due to a weak classifier being induced from the training data and propagating errors through each iteration. There are strategies, such as, guided self-training that attempts to eliminate these high-confidence errors (Drury et al. 2011).

[2.4] *Prediction of Future Information from Texts*

This area of related work concentrates upon work that uses past information in text to predict the likelihood of a future event. Radinsky & Horvitz (2013) used causal chains and probabilistic models to infer the likelihood of a specific event occurring in the future based upon current information. Hashimoto et al. (2014) used a supervised approach to learn causal chains and predict future events. They assumed that causality can be based on three assumptions: 1. two nouns that are joined by a binary semantic relation form causality between two events when combined with two predicates, 2. there are specific grammatical scenarios where causality will occur and 3. cause and events are strongly associated. Radinsky & Horvitz (2013) produced an algorithm called "Pundit" that generated event sce-

[3] A common alternative strategy is to propagate label from labelled to unlabelled instances in a transductive strategy (Rossi et al. 2014).

narios from a causal event. Kunneman & Van den Bosch (2012) used Tweets about Dutch football to predict future transfers of players.

[3] CORPUS

The corpus that we used for the experiments was news stories about agricultural in Brazil. These stories were gathered from various sources from the Internet from 1995 until 2014. The data was not contiguous, and consequently there were temporal gaps in the data. The stories were split into sentences and POS tagged with the De Alencar (2010). The corpus contained 295,307 sentences.

[3.1] *Manually Labelled Data*

Labelled data was required for the causal relation extraction and the sentiment classification tasks. A random set of 394 sentences were selected from the corpus. The data was categorized by a single annotator into two categories: causative and non-causative. The non-causative category had 84 sentences and the causative category had 310 sentences. The sentences in the causative category had one of the following categories added to their words: cause, effect, causative link or non-causative. The density of causative relations was high when compared to other causative relations annotation exercises we have undertaken (Drury et al. 2014a). This may be due to the type of text annoatated or the selection of sentences may have been atypical.

The labelled causative data was sub-divided into three categories (neutral, negative or positive) for the sentiment classification evaluation. The negative category had 228 sentences, the neutral 37 and the positive 45 sentences. The negative category was the majority class. This was unsurprising as most of the agricultural news stories were negative. Examples of the labelled data can be found in Table 1. The training data is available from `http://goo.gl/IYP1t1`.[4]

Category	Sentence
Negative	Recentemente, foram as geadas que afetaram os canaviais.
Negative	Fmc lança portal de informações sobre nematóides, praga que ameaça a cana de açúcar
Positive	o mercado internacional provocaram uma ligeira alta em o pregao de ontem

TABLE 1: Example of causative labelled data.

[4] The annotation schema for the data is: NC = non-causitive, CN = Cause Noun, EN = Effect Noun and CV = Causal Verb.

[4] ALGORITHM DESCRIPTION

The algorithm was designed to: 1. extract causal relations from text, 2. label cause, effect and casual link of the relation and 3. classify the causal relation into negative, neutral or positive categories.

[4.1] *Causal Relation Extraction*

The causal relation extraction (CRE) part of the algorithm is a multi-view self-training algorithm (Ando & Zhang 2007), that uses global and local classifiers to mitigate error propagation through the training iterations. This subsection will discuss in detail the CRE part of the algorithm and the motivation behind the choices made.

The global classifier is a relative link density (RLD) classifier (Drury et al. 2014c) that labels causative verbs in a sentence.[5] It is based upon a graph based approach that propagates causative and non-causative labels from labelled verbs to unlabelled verbs depending upon the link density between the verbs in the graph. The technique is described in full by (Drury et al. 2014c). RLD is complemented by a rule tagger that annotates noun phrases in sentences. The rule classifier is based upon a number of manually created decision rules. This combination of RLD and rule labeller attempts to identify the $NP\ V\ NP$ pattern described in the related work.

The local classifier is a combination stacked of CRFs. Stacking is a meta-learning technique where the training data is divided randomly between the CRFs. Each CRF produces a model, the models are used in combination to label casual relations in text. Each CRF has a "separate view" of the data, and consequently the number of errors produced by the models is reduced (Vilalta & Drissi 2002). The CRFs classify each word in a sentence as either: 1. Non-causative, 2. Causative Link, 3. Cause or 4. Effect. Classification sequences that match the aforementioned $NP\ V\ NP$ are assumed to be causal relations. There were two steps to train the CRFs. The steps were: feature selection and selection of the meta-learning technique.

Feature selection was achieved using a genetic algorithm (GA) (Nongmeikapam & Bandyopadhyay 2011) because: 1. it was not clear what the best features were and 2. the feature space was large, and it was not possible to test every feature combination. The GA used a pool of a 499 random solutions and 1 seed solution that contained all 54 categories of possible features. The GA used an accuracy figure from a hold-out evaluation as a fitness function. The hold-out evaluation used the manually labelled data described on section [3.1]. The hold-out evaluation ignored correct classifications for non-causative words because this class

[5] A list of causative verbs generated by a previous version of this algorithm is freely available from the resources described by (Drury et al. 2014b).

was the majority class and simply guessing this class for all words would have produced an accuracy of approximately 90.00% without correctly identifying any causal relations. The accuracy figure was calculated by the number of: 1. effect words, 2. causative link and 3. cause words classified correctly minus the number incorrect classification of non-causative and causative elements. The equation for the hold-out function is $\frac{Ccr}{Tcr+Enc}$, where Ccr is the number of correct causal relation elements classified (cause, effect, causal link), Tcr is the total number of causal relation elements and Enc is the number of erroneous classifications of non-causal words as a causal relation element.

The solutions were ranked by accuracy and the bottom 50% of the solutions were removed. The breeding strategy selected one surviving solution and chose randomly another surviving solution to breed with. The order of the features of the breeding solutions was randomized, and 50% of each solution was selected for the new solution. Duplicate features were removed. The mutation rate was 0.1, meaning that 25 of the new solutions were mutated. The mutation strategy took one feature of the solution and either: changed its value or swapped it for a new feature. The GA ran for 35 generations. The GA was limited to 35 generations because the GA was a time intensive process. The results are displayed in Figure 1. The diagram shows a steady increase over increasing generations with a number of plateaus. We hypothesize that the plateaus were caused by delay in the best solutions influencing the populations. The results represent a 14.28% relative increase over the initial "best solution" selected on the first generation. The results were unimpressive because 1. we excluded correct non-causative classifications from the fitness measure and 2. the limited amount of labelled data produced weak models.

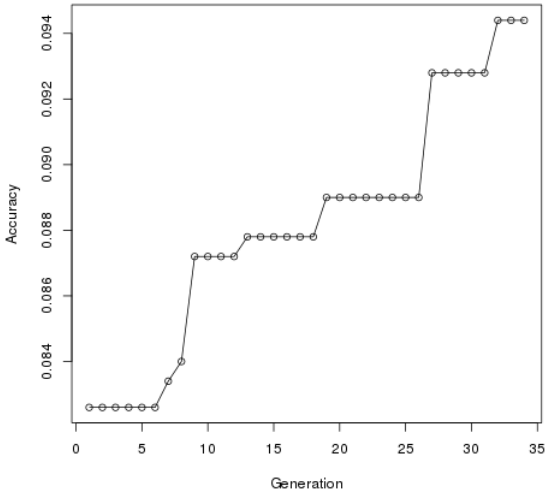

FIGURE 1: Evolution of accuracy with a GA feature selection

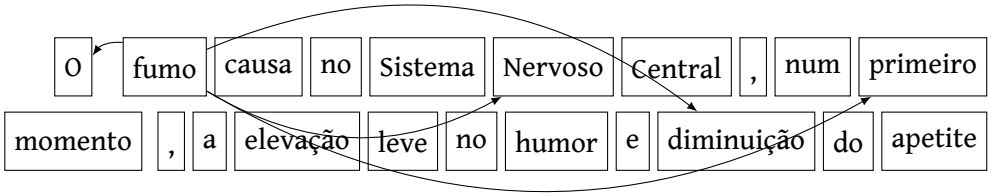

FIGURE 2: Examples of Word Dependencies in a Causal Relation for the Cause Candidate "fumo".

The categories of features selected by the GA strategy where: words ahead (number of words ahead) 16, 4, 8, word behind (number of words behind) 1, word features: number, punctuation, start of sentence, sentiment value, stopword and current word. An example of the features is provided in Figure 2, where the word features are demonstrated for the cause candidate "fumo". The "look behind" word is "O" and the "look ahead" words are: do, momento, Nervoso. Each of these words had a number of word specific features. For example, the cause candidate, "fumo", would have the following word features: IsStartOfSentence: false, Ispunctuation: false, HasSentimentValue: false, IsStopword:false and CurrentWord: fumo. Each of the look-ahead and look-behind word-features would be included in the features for the cause-candidate, "fumo".

In addition to using feature selection to improve the performance of the CRF we evaluated the effectiveness of meta-learning. The meta-learning technique we evaluated was stacking (Klugl et al. 2012) because the research literature suggests that stacking CRFs outperform a single CRF. The stacking strategy we attempted was to provide a separate random part of the training data to each individual CRF. The CRFs then vote on each classification with the majority vote being accepted as the classification of the stacked CRF.

We performed a basic evaluation of stacked 3 and 5 CRFs against a baseline of 1 CRF. The evaluation was a hold-out evaluation using he manually labelled data described on section [3.1]. The hold-out evaluation was 80:20 1 X 10 , where the data was randomly separated into two partitions: 80% for training and 20% for evaluation. The process was repeated 10 times. An average accuracy was calculated. We found that a stacked 3 CRFs performed gained the highest accuracy on the hold-out evaluation. A more in-depth evaluation was made that we describe later on in the paper.

[4.2] *Self-training*

The labelled data described on section [3.1] was limited, and consequently any model produced from this data would likely to be weak and produce errors. This

Name of Strategy	Accuracy Classification	Accuracy Annotation
Relative Link Classifier + Rule Labeller + Stacked CRF	0.81 $\pm$0.09	0.67 $\pm$0.09
Relative Link Classifier + Rule Labeller	0.61 $\pm$0.09	0.64 $\pm$0.09
Relative Link Classifier + Rule Labeller + Single CRF	0.76 $\pm$0.09	0.72 $\pm$0.09
Single CRF	0.13 $\pm$0.09	0.00 $\pm$0.00

TABLE 2: Analysis of Causal Relation Strategies.

characteristic of a weak classifier was shown in the feature selection experiments where the single classifier gained relatively low accuracy measures. A semi-supervised learning strategy is a method that combines labelled and unlabelled data to improve the performance of a classifier.

We choose self-training, that is an iterative technique that adds high confidence classifications of unlabelled data as training data in the next cycle. A weakness of self-training is error propagation where the classifier makes an error in classification that is then added to the training data that influences the next cycle. It is possible that classifier could have less accuracy after self-training than the model induced from the training data (Drury et al. 2011).

As stated earlier this algorithm used local and global classifiers to mitigate error propagation. We performed a number of experiments with various configurations of classifiers to supplement the limited hold-out evaluation we performed earlier. The experiments with self-training were designed to justify the selections made for the algorithm. The experiments allowed each configuration of classifiers to classify the whole corpus, and a random selection of 100 classifications were analyzed manually to produce an accuracy figure for: annotations and sentence classification. There was only one iteration for each classifier due to time constraints. The combinations analyzed were: 1. single Conditional Random Field, 2. Relative Link Classifier and Rule Labeller, 3. Relative Link Classifier and Rule Labeller with single Conditional Random Field, and 4. Relative Link Classifier and Rule Labeller with single Conditional Random Field. We calculated an error bar for that was based upon a confidence interval of 95%. The results are displayed in Table 2.

The results show that the combinations of the rule classifier with various combinations of CRFs out-performed: Relative Link Classifier and a Single Conditional Random Field . The stacked CRF was the only combination that outperformed the Relative Link Classifier by more than the margin of error, consequently it was chosen for the causative relation extraction of our algorithm. The relative poor performance of the CRF reflected our experience in the feature selection phase. The causal relation extraction self-training algorithm is fully described in Algorithm 1.

Input: UL,LD, DR
Output: LD
/* UL = unlabelled data, LD = labelled data, DR = decision rules */
while *True* **do**
  $gc \leftarrow train(LD)$;
  $crf \leftarrow train(LD)$;
  /* gc = RLC, crf = Conditional Random Fields (stacked) */
  $count \leftarrow 0$;
  for $sentence \in UL$ **do**
    /* test if sentence is in labelled data */
    if $sentence$ *in* LD **then**
      $continue$;
    end
    /* test agreement for verbs v, cause c and effect e */
    $e, c, v = classify(DR, gc, sentence)$;
    $e_1, c_1, v_1 = classify(crf, sentence)$;
    if $e == e_1$ *and* $c == c_1$ *and* $v == v_1$ **then**
      $count \leftarrow count + 1$;
      /* Add training candidate to labelled data */
      $LD \leftarrow appendData(LD, e, c, v)$;
    end
  end
  /* Termination Condition */
  if $count == 0$ **then**
    return LD;
  end
end

Algorithm 1: Self-training algorithm

[4.3] *Sentiment Classification*

The second part of the algorithm classifies causal relations extracted by the first part of the algorithm into one of three sentiment categories (positive, negative or neutral). The algorithm achieves this by: removing the cause part of the causative relation, and classifying the remaining part of the relation into one of aforementioned categories.

The sentiment classification part of the algorithm is the Guided Self Training algorithm described by Drury et al. (2011) who used a combination of rules and self-training to produce a "strong" classifier. This strategy has two parts: dictionary construction and self-training.

Positive	Negative
avança, atraente, boas, elevar, belo	prejuízos, baixa, danos, perdas
benévolo, favorável, ótimo, benigno	geadas, quebra, diminuição, falta

TABLE 3: Examples of sentiment words from the dictionary construction process.

Dictionary construction was achieved by extracting: adjectives, adverbs and nouns from the training data. These words are expanded with synonyms from Onto.pt (Gonçalo Oliveira 2014). Onto.pt is a taxonomy of Portuguese words that are organized by synsets of related words. The synonyms were extracted by: 1. loading the taxonomy into the rdflib python library[6] and 2. returning words (synonyms) from the same synset as a target word.

The training data was constructed by dividing the training data described on section [3] into three sentiment categories: neutral, negative and positive. This data was used for dictionary construction and as training data for a classifier. The positive dictionary had 312 entities, where as the negative dictionary had 4767 entries. This indicates that the training data was overwhelmingly negative. An example of the entries are described in Table 3.

The linguistic rules are the rules described by Drury et al. (2011) where a causal relation is classified in one of the sentiment classes with the following criteria: 1. a sentence is classified as positive if it has two or more entries from the positive class and none from the negative dictionary, 2. a sentence is classified as negative if it has two or more entries from the negative dictionary and none from the positive dictionary, 3. a sentence is classified neutral if it contains no entries from either the positive or negative dictionaries, and 4. if a sentence contains one entry from the positive or the negative dictionaries then no classification is made.

The guided self-training strategy was adjusted to use balancing strategies to improve the performance of the induced model. We used random over balancing that has been shown to gain good results in sentiment classification of Portuguese (Drury & Lopes 2014). The guided self-training algorithm for sentiment classification is described in Algorithm 2.

Guided Self-training evaluation:

The suitability of the sentiment classification strategy was evaluated with 80:20 1 X 10 hold-out evaluation. The hold-out evaluation relied upon labelled data, that in this case was the labelled sentiment data described on section [3.1]. The hold-out evaluation reversed 80% of the data for training and 20% for testing. The test was repeated 10 times with different splits of the data. The competing strategies

[6] http://code.google.com/p/rdflib/.

Input: UL,LD, DR, MC
Output: SC
/* UL = unlabelled data, LD = labelled data, DR = decision rules, Minimum Confidence, SC = Sentiment Classifier */
while *True* **do**
  /* Balance Training Data */
  $LD1 = Balance(LD)$;
  $sc \leftarrow train(LD1)$;
  /* sc = sentiment classifier */
  $count \leftarrow 0$;
  for $sentence \in UL$ **do**
    /* test if sentence is in labelled data */
    if $sentence$ *in* LD **then**
      $continue$;
    end
    $DRc = classify(DR, sentence)$;
    $scc = classify(sc, sentence, MC)$;
    if $scc == None$ **then**
      $continue$;
    end
    /* Add training candidate to labelled data */
    $count = count + 1$;
    if $Drc == None$ *or* $scc == DRc$ **then**
      $LD \leftarrow appendData(LD, scc, sentence)$;
    else
      $LD \leftarrow appendData(LD, DRc, sentence)$;
    end
  end
  /* Termination Condition */
  if $count == 0$ **then**
    return sc;
  end
end

Algorithm 2: Guided Self-training.

were tested on the same splits. The evaluation measure was accuracy. The results are displayed in Table 4. The results clearly show that the guided self-training strategy produced the superior results.

Strategy	Accuracy
Supervised	0.73 ± 0.04
Guided Self-Training	0.84 ± 0.06

TABLE 4: Results for Hold-Out Evaluation.

[5] SENTIMENT PREDICTION

The last step in the strategy is to assign a sentiment probability to a cause. This is achieved by grouping common causes and aggregating their sentiment categories to produce a sentiment distribution for a specific cause. This grouping process is illustrated in the following example. We have three causative sentences and their sentiment categories: 1. "chuva causa cheias no Porto", neutral, 2. "chuva causa danos em Minas Gerais", negative and 3. "Chuva causa inundações e destrói casa em Itapetininga", negative. When the cause is "chuva", and its sentiment distribution would be $P = \{Neu = 0.33, Neg = 0.66, Pos = 0.0\}$.

[5.1] *Experiments*

The experiments for sentiment prediction manually evaluated the sentiment classifications for specific common causes. In the experiments we ran the aforementioned causal relation extractor and sentiment classifier. The relations were grouped by cause and their sentiment distributions calculated. There were 4988 common causes. The most frequent sentiment causal events and their sentiment distributions are displayed in Table 5.

No. Causal Rel.	Cause Event	Sent Dist.
116	seca	neg 0.66 pos 0.05 neu 0.28
95	estiagem	neg 0.58 pos 0.13 neu 0.29
76	chuvas	neg 0.41 pos 0.04 neu 0.55
73	cana açúcar	neg 0.16 pos 0.1 neu 0.74
70	chuva	neg 0.36 pos 0.01 neu 0.63
59	clima	neg 0.56 pos 0.12 neu 0.32
41	governo	neg 0.07 pos 0.17 neu 0.76
38	brasil	neg 0.13 pos 0.18 neu 0.68
35	crise	neg 0.63 pos 0.06 neu 0.31
30	cana	neg 0.13 pos 0.27 neu 0.6

TABLE 5: Frequent Causal Events and their Sentiment Distribution.

Cause Event	Acc. Sentiment Category	Acc. Causal Relation
Expansão	0.83	1.0
Pessoas	0.29	1.0
Petrobras	1.0	1.0
Baixas Temperaturas	0.67	1.0
Praga	1.0	1.0
Homen	0.54	1.0
Canais	1.0	1.0
Conab	0.69	0.31
Praticidade	1.0	1.0
Aquecimento Global	0.67	1.0

TABLE 6: Accuracy for Causal Events.

[5.2] *Evaluation*

We performed a manual evaluation where we randomly selected 10 cause event groups and evaluated the causal relations that constitute the sentiment distribution. The evaluation tested if: the sentiment category was correct and it was a causal relation.

The causal events chosen were: expansão, pessoas, petrobras, baixas temperaturas geadas, praga, homem, canais, conab, praticidade and aquecimento global. The results are shown in Table 6.

The accuracy of the whole sample for: 1. causative relation detection was 0.91 and 2. sentiment classification was 0.77. We can therefore calculate the overall accuracy as 0.70 for extracting and classifying causal sentimental relations.

The causal relation extraction strategy performed poorly when the common cause event was Conab.[7] This was a special case because it is an organization that made: 1. predictions about future events or 2. showed possible effects from a cause. These statements had causal characteristics, but were not causal relations, for example, "Estudo da Conab mostra impacto do clima nas lavouras".

The errors made by the sentiment classification were between: 1. negative and neutral categories and 2. positive and neutral categories. This type of error is less serious than classifying a negative relation as positive or vice-versa because any inference based from this sentiment mistake will be ignored.

[6] CONCLUSION AND FUTURE WORK

This work introduces a new type of sentiment analysis where we predict a sentiment distribution from a cause event. The initial results are encouraging as they

[7] http://www.conab.gov.br.

seem to make "intuitive" sense. For example, "seca[8]" will be mainly negative for agriculture because of future lower crop yields, however it seems reasonable that there may be some positive future news (for farmers) in the form of crop price rises due to lower supply and constant demand, although this news could be seen as negative for consumers.

The future work is to evaluate the predictive ability of sentiment distributions of causes. This work is centred around agriculture, and causes such as "falta de chuva" or "seca" are likely to have similar effects on crops in the future as they have had in the past. It is reasonable to assume at least in this domain that we can estimate the sentiment distribution of future news stories. This may allow the improvement of time dependent sentiment tasks such as reputation management and stock trading.

ACKNOWLEDGEMENTS

This work was supported by FAPESP grant number: 11/20451-1.

REFERENCES

Altenberg, Bengt. 1984. Causal linking in spoken and written english. *Studia Linguistica* 38(1). 20–69.

Ando, Rie Kubota & Tong Zhang. 2007. Two-view feature generation model for semi-supervised learning. In *Proceedings of the 24th international conference on machine learning*, 25–32. ACM.

Baron, Naomi S. 1974. The structure of english causatives. *Lingua* 33(4). 299–342.

De Alencar, Leonel Figueiredo. 2010. Uma ferramenta para anotação automática de corpora usando o NLTK. In *The 9th brazilian corpus linguistics meeting*, s/pp.

Drury, Brett, Paula C. F. Cardoso, Jorge Carlos Valverde-Rebaza, Alan Valejo, Fabio Pereira & Alneu de Andrade Lopes. 2014a. An open source tool for crowdsourcing the manual annotation of texts. In *Computational processing of the portuguese language - 11th international conference, PROPOR*, 268–273.

Drury, Brett, Paula C.F. Cardoso, Janie M. Thomas & Alneu de Andrade Lopes. 2014b. Lexical resources for the identification of causative relations in Portuguese texts. In *Proceedings of workshop on tools and resources for automatically processing Portuguese and Spanish*, s/pp.

Drury, Brett & Alneu Lopes. 2014. A comparison of the effect of feature selection and balancing strategies upon the sentiment classification of Portuguese news stories. In *Proceedings of ENIAC*, s/pp.

[8] Table 5.

Drury, Brett, Rafael Geraldeli Rossi & Alneu de Andrade Lopes. 2014c. Identification of Brazilian Portuguese causative verbs through a weighted graph classification strategy. In *Computational Processing of the Portuguese Language*, 274–279. Springer.

Drury, Brett, Luís Torgo & J. J Almeida. 2011. Guided self training for sentiment classification. In *Proceedings of robust unsupervised and semi-supervised methods in natural language processing workshop, RANLP conference*, 9–16. ACL.

Gonçalo Oliveira, Hugo. 2014. The creation of Onto.PT: A wordnet-like lexical ontology for Portuguese. In *Proceedings of computational processing of the portuguese language - 11th international conference (propor 2014)*, vol. 8775, 161–169. Springer.

Hashimoto, Chikara, Kentaro Torisawa, Julien Kloetzer, Motoki Sano, István Varga, Jong-Hoon Oh & Yutaka Kidawara. 2014. Toward Future Scenario Generation: Extracting Event Causality Exploiting Semantic Relation, Context, and Association Features. In *Proceedings of the 52nd annual meeting of the association for computational linguistics*, vol. 1, 987–997.

He, Yulan & Deyu Zhou. 2011. Self-training from labelled features for sentiment analysis. *Information Processing & Management* 47(4). 606–616.

Joskowicz, L., T. Ksiezyck & R. Grishman. 1989. Deep domain models for discourse analysis. In *Proceedings of the Annual AI Systems in Government Conference*, 195–200.

Kaplan, Randy. 1991. Knowledge-based acquisition of causal relationships in text. *Knowledge Acquisition* 3(3). 317–337.

Klugl, Peter, Martin Toepfer, Florian Lemmerich, Andreas Hotho & Frank Puppe. 2012. Stacked conditional random fields exploiting structural consistencies. In Pedro Latorre Carmona, J. Salvador Sánchez & Ana Fred (eds.), *Proceedings of 1st international conference on pattern recognition applications and methods ICPRAM*, 240–248. SciTePress.

Kunneman, F. & A. Van den Bosch. 2012. Leveraging unscheduled event prediction through mining scheduled event tweets. In N. Roos, M. Winands & J. Uiterwijk (eds.), *Proceedings of the 24th Benelux Conference on Artficial Intelligence*, 147.

Levin, Beth. 1993. *English verb classes and alternations*. University of Chicago Press.

Mehrabi, S., A. Krishnan, E. Tinsley, J. Sligh, N. Crohn, H. Bush, J. Depasquale, J. Bandos & M. Palakal. 2013. Event causality identification using conditional random field in the geriatric care domain. In *Proceedings of the 12th International Conference on Machine Learning and Applications*, vol. 1, 339–343.

Nongmeikapam, Kishorjit & Sivaji Bandyopadhyay. 2011. Genetic algorithm (GA) in feature selection for CRF based manipuri multiword expression (MWE) identification. *International Journal of Computer Science & Information Technology* 3(5). 53–66.

Pang, Bo, Lillian Lee & Shivakumar Vaithyanathan. 2002. Thumbs Up?: Sentiment Classification Using Machine Learning Techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, vol. 10, 79–86.

Radinsky, Kira & Eric Horvitz. 2013. Mining the web to predict future events. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, 255–264.

Riaz, Mehwish & Roxana Girju. 2014. Recognizing causality in verb-noun pairs via noun and verb semantics. In *Proceedings of the Workshop on Computational Approaches to Causality in Language (EACL)*, 48–57. The Association for Computer Linguistics.

Rossi, Rafael G., Alneu A. Lopes & Solange O. Rezende. 2014. A parameter-free label propagation algorithm using bipartite heterogeneous networks for text classification. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, 79–84. ACM.

Vilalta, Ricardo & Youssef Drissi. 2002. A perspective view and survey of meta-learning. *Artificial Intelligence Review* 18. 77–95.

CONTACTS

Brett Drury
Universidade de São Paulo
Brett.Drury@gmail.com

Alneu de Andrade Lopes
Universidade de São Paulo
alneu@icmc.usp.br