

Recognizing Causality in Verb-Noun Pairs via Noun and Verb Semantics

Mehwish Riaz and Roxana Girju

Department of Computer Science and Beckman Institute
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
{mriaz2, girju}@illinois.edu

Abstract

Several supervised approaches have been proposed for causality identification by relying on shallow linguistic features. However, such features do not lead to improved performance. Therefore, novel sources of knowledge are required to achieve progress on this problem. In this paper, we propose a model for the recognition of causality in verb-noun pairs by employing additional types of knowledge along with linguistic features. In particular, we focus on identifying and employing semantic classes of nouns and verbs with high tendency to encode cause or non-cause relations. Our model incorporates the information about these classes to minimize errors in predictions made by a basic supervised classifier relying merely on shallow linguistic features. As compared with this basic classifier our model achieves 14.74% (29.57%) improvement in F-score (accuracy), respectively.

1 Introduction

The automatic detection of causal relations is important for various natural language processing applications such as question answering, text summarization, text understanding and event prediction. Causality can be expressed using various natural language constructions (Girju and Moldovan, 2002; Chang and Choi, 2006). Consider the following examples where causal relations are encoded using (1) a verb-verb pair, (2) a noun-noun pair and (3) a verb-noun pair.

1. Five shoppers were **killed** when a car **blew up** at an outdoor market.
2. The attack on Kirkuk's police intelligence complex sees further **deaths** after **violence** spilled over a nearby shopping mall.

3. At least 1,833 people **died** in **hurricane**.

Since, the task of automatic recognition of causality is quite challenging, researchers have addressed this problem by considering specific constructions. For example, various models have been proposed to identify causation between verbs (Bethard and Martin, 2008; Beamer and Girju, 2009; Riaz and Girju, 2010; Do et al., 2011; Riaz and Girju, 2013) and between nouns (Girju and Moldovan, 2002; Girju, 2003). Do et al. (2011) have worked with verb-noun pairs for causality detection but they focused only on a small list of predefined nouns representing events.

In this paper, we focus on the task of identifying causality encoded by verb-noun pairs (example 3). We propose a novel model which first predicts cause or non-cause relations using a supervised classifier and then incorporates additional types of knowledge to reduce errors in predictions. Using a supervised classifier, our model identifies causation by employing shallow linguistic features (e.g., lemmas of verb and noun, words between verb and noun). Such features have been used successfully for various NLP tasks (e.g., part-of-speech tagging, named entity recognition, etc.) but confinement to such features does not help much to achieve performance for identifying causation (Riaz and Girju, 2013). Therefore, in our model we plug in additional types of knowledge to obtain better predictions for the current task. For example, we identify the semantic classes of nouns and verbs with high tendency to encode cause or non-causal relations and use this knowledge to achieve better performance. Specifically, the contributions of this paper are as follows:

- In order to build a supervised classifier, we use the annotations of FrameNet to generate a training corpus of verb-noun instances encoding cause and non-cause relations. We propose a set of linguistic features to learn and identify causal relations.

- In order to make intelligent predictions, it is important for our model to have knowledge about the semantic classes of nouns with high tendency to encode causal or non-causal relations. For example, a named entity such as person, organization or location may have high tendency to encode non-causality unless a metonymic reading is associated with it. In our approach, we identify such semantic classes of nouns by exploiting a named entity recognizer, the annotations of frame elements provided in FrameNet and WordNet.
- Verbs are the important components of language for expressing events of various types. For example, Pustejovsky et al. (2003) have classified events into eight semantic classes: OCCURRENCE, PERCEPTION, REPORTING, ASPECTUAL, STATE, I_STATE, L_ACTION and MODAL. We argue that there are some semantic classes in this list with high tendency to encode cause or non-cause relations. For example, reporting events represented by verbs say, tell, etc., have high tendency to just report other events instead of encoding causality with them. In our model, we use such information to reduce errors in predictions.
- Each causal relation is characterized by two roles i.e., cause and its effect. In example 3 above, the noun “hurricane” is cause and the verb “died” is its effect. However, a verb-noun pair may not encode causality when a verb and a noun represent same event. For example, in instance “Colin Powell **presented** further evidence in his **presentation**.”, the verb “presented” and the noun “presentation” represent same event of “presenting” and thus encoding non-cause relation with each other. In our model, we determine the verb-noun pairs representing same or distinct events to make predictions accordingly.
- We adopt the framework of Integer Linear Programming (ILP) (Roth and Yih, 2004; Do et al., 2011) to combine all the above types of knowledge for the current task.

This paper is organized as follows. In next section, we briefly review the previous research done for identifying causality. We introduce our model and evaluation with discussion on results in section 3 and 4, respectively. The section 5 of the paper concludes our current research.

2 Related Work

In computational linguistics, researchers have always shown interest in the task of automatic recognition of causal relations because success on this task is critical for various natural language applications (Girju, 2003; Chklovski and Pantel, 2004; Radinsky and Horvitz, 2013).

Following the successful employment of linguistic features for various tasks (e.g., part-of-speech tagging, named entity recognition, etc.), initially NLP researchers proposed approaches relying mainly on such features to identify causality (Girju, 2003; Bethard and Martin, 2008; Sporleder and Lascarides, 2008; Pitler and Nenkova, 2009; Pitler et al., 2009). However, researchers have recently shifted their attention from these features and tried to consider other sources of knowledge for extracting causal relations (Beamer and Girju, 2009; Riaz and Girju, 2010; Do et al., 2011; Riaz and Girju, 2013). For example, Riaz and Girju (2010) and Do et al. (2011) have proposed unsupervised metrics for learning causal dependencies between two events. Do et al. (2011) have also incorporated minimal supervision with unsupervised metrics. For a pair of events (a, b), their model makes the decision of cause or non-cause relation based on unsupervised co-occurrence counts and then improves this decision by using minimal supervision from the causal and non-causal discourse markers (e.g., because, although, etc.).

In search of novel and effective types of knowledge to identify causation between two verbal events, Riaz and Girju (2013) have proposed a model to learn a Knowledge Base (KB_c) of verb-verb pairs. In this knowledge base, the English language verb-verb pairs are automatically classified into three categories: (1) Strongly Causal, (2) Ambiguous and (3) Strongly Non-Causal. The Strongly Causal and Strongly Non-Causal categories contain verb-verb pairs with highest and least tendency to encode causality, respectively and rest of the verb-verb pairs are considered ambiguous with tendency to encode both types of relations. They claim that this knowledge base of verb-verb pairs is a rich source of causal associations. The incorporation of this resource into a causality detection model can help identifying causality with better performance. In this research, we also try to go beyond the scope of shallow linguistic features and identify additional

interesting types of knowledge for the current task.

3 Computational Model for Identifying Causality

In this section, we introduce our model for identifying causality encoded by verb-noun pairs. Specifically, we extract all main verbs and noun phrases from a sentence and predict cause or non-cause relation on verb-noun_phrase (v-np) pairs. In order to make task easier, we consider only those v-np pairs where v (verb) is grammatically connected to np (noun phrase). We assume that a v and np are grammatically connected if there exists a dependency relation between them in the dependency tree. We apply a dependency parser (Marnette et al., 2006) to identify such dependencies. Our model first employs a supervised classifier relying on linguistic features to make binary predictions (i.e., does a verb-noun_phrase pair encode a cause or non-cause relation?). We then incorporate additional types of knowledge on top of these binary predictions to improve performance.

3.1 Supervised Classifier

In this section, we propose a basic supervised classifier to identify causation encoded by v-np pairs. To set up this supervised classifier, we need a training corpus of instances of v-np pairs encoding cause and non-cause relations. For this purpose, we employ the annotations of FrameNet project (Baker et al., 1998) provided for verbs. For example, consider the following annotation from FrameNet for the verb “dying” with argument “solvent abuse” where the pair “dying-solvent abuse” encodes causality.

A campaign has started to try to cut the rising number of children **dying** [*cause* from **solvent abuse**].

To generate a training corpus, we collect annotations of verbs from FrameNet s.t. the annotated element (aka. frame element) is a noun phrase. For example, we get a causal training instance of “dying-solvent abuse” pair from the above annotation. We assume that if a FrameNet’s annotated element contains a verb in it then this may not represent a training instance of v-np pair. For example, we do not consider the following annotation in our training corpus where causality is encoded between two verbs i.e., “died-fell”.

A fitness fanatic **died** [*cause* when 26 stone of weights **fell** on him as he exercised].

After extracting training instances from FrameNet, we assign them cause (c) and non-cause ($\neg c$) labels. We manually examined the inventory of labels of FrameNet and use the following scheme to assign the c or $\neg c$ to each training instance. All the annotations of FrameNet with following labels are considered as causal training instances and rest of the annotations are considered as non-causal training instances.

Purpose, Internal cause, Result, External cause, Cause, Reason, Explanation, Required situation, Purpose of Event, Negative consequences, resulting action, Effect, Cause of shine, Purpose of Goods, Response action, Enabled situation, Grinding cause, Trigger

For this work, we have acquired 2,158 (65,777) cause (non-cause) training instances from FrameNet. Since, the non-cause instances are very large in number, our supervised model tends to assign non-cause labels to almost all instances. Therefore, we employ equal number of cause and non-cause instances for training. In future, we plan to extract more annotations from the FrameNet and employ more than one human annotators to assign the labels of cause and non-cause relations to the full inventory of labels of FrameNet.

- **Lexical Features:** 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.
- **Semantic Features:** We adopted this feature from Girju (2003) to capture the semantics of nouns. The 9 noun hierarchies of WordNet i.e., entity, psychological feature, abstraction, state, event, act, group, possession, phenomenon are used as this feature. Each of these hierarchies is set to 1 if any sense of the head noun of noun phrase lies in that hierarchy otherwise set to 0.
- **Structural Features:** This feature is applied by considering both subject (i.e., sub_in_np) and object (i.e., obj_in_np) of a verb. For example, for a v-np pair the variable sub_in_np is set to 1 if the subject of v is contained in np, set to 0 if the subject of v is not contained in np and set to -1 if the subject of v is not available in the instance. The subject and object of a verb are its core arguments and may sometime be part of an event represented by a verb. Therefore, these argument may have high tendency to encode non-cause relations.

We set up the following integer linear program after acquiring predictions of c and $\neg c$ labels using our supervised classifier.

$$Z_1 = \max \sum_{v\text{-np} \in I} \sum_{l \in L_1} x_1(v\text{-np}, l) P(v\text{-np}, l) \quad (1)$$

$$\sum_{l \in L_1} x_1(v\text{-np}, l) = 1 \quad \forall v\text{-np} \in I \quad (2)$$

$$x_1(v\text{-np}, l) \in \{0, 1\} \quad \forall v\text{-np} \in I \quad \forall l \in L_1 \quad (3)$$

Here $L_1 = \{c, \neg c\}$, I is the set of all instances of $v\text{-np}$ pairs and $x_1(v\text{-np}, l)$ is the decision variable set to 1 only if the label $l \in L_1$ is assigned to $v\text{-np}$. The Equation 2 constraints that only one label out of $|L_1|$ choices can be assigned to a $v\text{-np}$ pair. The equation 3 requires $x_1(v\text{-np}, l)$ to be a binary variable. Specifically, we try to maximize the objective function Z_1 (equation 1) which assigns the label cause or non-cause to all $v\text{-np}$ pairs (i.e., set the variables $x_1(v\text{-np}, l)$ to 1 or 0 for all $l \in L_1$ and for all $v\text{-np}$ pairs in I) depending on the probabilities of assignment of labels (i.e., $P(v\text{-np}, l)$)¹. These probabilities can be obtained by running a supervised classification algorithm (e.g., Naive Bayes and Maximum Entropy). In our experiments, we provide results using the following probabilities acquired with Naive Bayes.

$$\begin{aligned} P(v\text{-np}, c) &= 1.0 - \frac{\sum_{k=1}^n \log P(f_k | c)}{\sum_{k=1}^n \sum_{l \in \{c, \neg c\}} \log P(f_k | l)} \\ P(v\text{-np}, \neg c) &= 1.0 - P((v, np), c) \end{aligned} \quad (4)$$

where f_k is a feature, n is total number of features and $P(f_k | l)$ is the smoothed probability of a feature f_k given the training instances of label l .

3.2 Knowledge of Semantic classes of nouns

Philosopher Jaegwon Kim (Kim, 1993) (as cited by Girju and Moldovan (2002)) pointed out that the entities which represent either causes or effects are often events, but also conditions, states, phenomena, processes, and sometimes even facts. Therefore, according to this our model should have knowledge of the semantic classes of noun phrases with high tendency to encode cause or non-cause relations. Considering this type of knowledge, we can automatically review and correct the wrong predictions made by our basic supervised classifier.

¹We use the integer linear program solver available at <http://sourceforge.net/projects/lpsolve/>

We argue that if a noun phrase represents a named entity then it can have least tendency to encode causal relations unless there is a metonymic reading associated with it. For example, consider the following cause and non-cause examples where noun phrase is a named entity.

4. Sandy **hit Cuba** as a Category 3 hurricane.
5. Almost all the weapon sites in Iraq were **destroyed** by the **United States**.

In example 4, Cuba is location and does not encode causality. However, in example 5 the pair “destroyed-the United States” encode causality where a metonymic reading is associated with the location. We apply Named Entity Recognizer (Finkel et al., 2005) and assume if a noun phrase is identified as a named entity then its corresponding verb-noun_phrase pair encodes non-cause relation. This constraint can lead to a false negative prediction when the metonymic reading is associated with a noun phrase. In order to avoid as much false negatives as possible, we imply the following simple rule i.e., if one of the following cue words appear between a verb and a noun phrase then do not apply the constraint stated above.

by, from, because of, through, for

In our experiments, the above simple rule helps avoiding some false negatives but in future any subsequent improvement with a better metonymy resolver (Markert and Nissim, 2009) should improve the performance of our model.

In addition to named entities, there can be various noun phrases with least tendency to encode causation. Consider the following example, where “city” is a location and does not encode cause-effect relation with the verb “remained”.

Substantially fewer people **remained** in the **city** during the Hurricane Ivan evacuation.

In this work, we identify the semantic classes of noun phrases which do not normally represent events, conditions, states, phenomena, processes and thus have high tendency to encode non-cause relations. For this purpose, we manually examine the inventory of labels assigned to noun phrases in FrameNet (see table 1) and classify these labels into two classes (c_{np} and $\neg c_{np}$). Here, the class c_{np} ($\neg c_{np}$) represents the labels of noun phrases with high (less) tendency to encode cause-effect relations. For example, the label “Place” $\in \neg c_{np}$

(see table 1) represents a location and it may have least tendency to encode causality if metonymy is not associated with it. Using the classification of frame elements in table 1, we obtain the annotations of noun phrases from FrameNet and categorize these annotations into c_{np} and $\neg c_{np}$ classes. On top of the annotations of these two semantic classes, we build a supervised classifier for predicting c_{np} or $\neg c_{np}$ label for the noun phrases. After obtaining predictions, we select all noun phrases lying in class $\neg c_{np}$ and apply the same constraint stated above for the named entities. We use the following set of features to set up a supervised classifier for c_{np} and $\neg c_{np}$ labels.

- **Lexical Features:** words of noun phrase, lemmas of all words of noun phrase, head word of noun phrase, first two (three) (four) letters of head noun of noun phrase, last two, (three) (four) letters of head noun of noun phrase.
- **Word Class Features:** part-of-speech tags of all words of noun phrase, part-of-speech tag of head noun of noun phrase.
- **Semantic Features:** all (frequent) sense(s) of head noun of noun phrase.

We have acquired 23,334 (81,279) training instances of c_{np} ($\neg c_{np}$) class, respectively for this work. We also use WordNet to obtain more training instances of these classes. We follow the approach similar to Girju and Moldovan (2002) and adopt some senses of WordNet (shown in table 1) to acquire training instances of noun phrases. For example, considering the table 1, we assign $\neg c_{np}$ label to any noun whose all senses in WordNet lie in the semantic hierarchy originated by the sense {time period, period of time, period}. Following this scheme, we extract instances of nouns and noun phrases from English GigaWord corpus and assign the labels c_{np} and $\neg c_{np}$ to them by employing WordNet senses given in table 1. Girju and Moldovan (2002) have used similar scheme to rank noun phrases according to their tendency to encode causation. In comparison to them, we use the WordNet senses to increase the size of our training set of noun phrases obtained using FrameNet above. In addition to this, we build a automatic classifier on the training data obtained using labels of FrameNet and WordNet senses to classify noun phrases of test instances into two semantics classes (i.e., c_{np} and $\neg c_{np}$). In our training corpus of there are 2, 214, 68 instances of noun phrases (50% belongs to each of c_{np} and $\neg c_{np}$

classes).

We incorporate the knowledge of semantics of nouns in our model by making the following additions to the integer linear program introduced in section 3.1.

$$Z_2 = Z_1 + \sum_{np:v-np \in I} \sum_{l \in L_2} x_2(np, l)P(np, l) \quad (5)$$

$$\sum_{l \in L_2} x_2(np, l) = 1 \quad \forall np : v-np \in I - M \quad (6)$$

$$x_2(np, l) \in \{0, 1\} \quad \forall np : v-np \in I - M \quad (7)$$

$$\forall l \in L_2$$

$$x_1(v-np, \neg c) - x_2(np, \neg c_{np}) \geq 0 \quad (8)$$

$$\forall np : v-np, \quad \forall v-np \in I - M$$

Here $L_2 = \{c_{np}, \neg c_{np}\}$ and M is the set of instances of those v-np pairs for which we consider the possibility of attachment of metonymic reading with np, $x_2(np, l)$ is the decision variable set to 1 only if the label $l \in L_2$ is assigned to np. The Equation 6 constraints that only one label out of $|L_2|$ choices can be assigned to a np. The equation 7 requires $x_2(np, l)$ to be a binary variable. The constraint 8 assumes that if an np belongs to the semantic class $\neg c_{np}$ then its corresponding pair v-np is assigned the label $\neg c$. We maximize the objective function Z_2 (equation 5) of our integer linear program subject to the constraints introduced above. We predict the semantic class of a noun phrase using the supervised classifier for c_{np} and $\neg c_{np}$ classes and set the probabilities i.e., $P(np, l) = 1, P(np, \{L_2\} - \{l\}) = 0$ if the label $l \in L_2$ is assigned to np. Again we use Naive Bayes to predict the labels for noun phrases. Also before running this supervised classifier, we run the named entity recognizer and assign $\neg c_{np}$ labels to all noun phrases identified as named entities. For our model, we apply named entity recognizer for seven classes i.e., LOCATION, PERSON, ORGANIZATION, DATE, TIME, MONEY, PERCENT (Finkel et al., 2005).

3.3 Knowledge of Semantic classes of verbs

In this section, we introduce our method to incorporate the knowledge of semantic classes of verbs to identify causation. Verbs are the components of language for expressing events of various types. In TimeBank corpus, Pustejovsky et al. (2003) have introduced eight semantic classes of events i.e., OCCURRENCE, PERCEPTION, REPORTING, ASPECTUAL, STATE, I.STATE, I.ACTION and MODAL. According to the definitions of these classes provided by Pustejovsky

Semantic Class	FrameNet Labels	WordNet Senses
c_{np}	Event, Goal, Purpose, Cause, Internal cause, External cause, Result, Means, Reason, Phenomena	{act, deed, human action, human activity}, {phenomenon}, {state}, {psychological feature}, {event}, {causal agent, cause, causal agency}
$\neg c_{np}$	Artist, Performer, Duration, Time, Place, Distributor, Area, Path, Direction, Sub-region	{time period, period of time, period}, {measure, quantity, amount}, {group, grouping}, {organization, organisation}, {time unit, unit of time}, {clock time, time}

Table 1: This table presents some examples of FrameNet labels in c_{np} and $\neg c_{np}$ classes. The full set of labels in both semantic classes are given in appendix A. It also presents the WordNet senses of nouns lying in c_{np} and $\neg c_{np}$ classes.

et al. (2003), the reporting events describe the action of a person, declare something or narrate an event e.g., the reporting events represented by verbs say, tell, etc. Here, we argue that a reporting event has the least tendency to encode causation because such an event only describes or narrates another event instead of encoding causality with it. We assume that the verbs representing reporting events have least tendency to encode causation and thus their corresponding v-np pairs have least tendency to encode causation. To add this knowledge to our model, we consider two classes of verbs i.e., c_v and $\neg c_v$ where the class c_v ($\neg c_v$) contains the verbs with high (less) tendency to encode causation. Using above argument we claim that all verbs representing reporting events belong to $\neg c_v$ class and verbs representing rest of the types of events belong to c_v class. We build a supervised classifier which automatically classifies verbs into c_v and $\neg c_v$ classes. We extract the instances of verbal events (i.e., verbs or verbal phrases) from TimeBank corpus and assign the labels c_v and $\neg c_v$ to these instances. Using these labeled instances, we build a supervised classifier by adopting the same set features as introduced in Bethard and Martin (2006) to identify semantic classes of verbs. Due to space constraint, we refer the reader to Bethard and Martin (2006) for the details of features. Again we use Naive Bayes to take predictions of c_v and $\neg c_v$ labels and their corresponding probabilities using equation 4.

We incorporate the knowledge of semantics of verbs in our model by making the following additions to the integer linear program.

$$Z_3 = Z_2 + \sum_{v: v\text{-np} \in I} \sum_{l \in L_3} x_3(v, l) P(v, l) \quad (9)$$

$$\sum_{l \in L_3} x_3(v, l) = 1 \quad \forall v : v\text{-np} \in I \quad (10)$$

$$x_3(v, l) \in \{0, 1\} \quad \forall v : v\text{-np} \in I \quad \forall l \in L_3 \quad (11)$$

$$x_1(v\text{-np}, \neg c) - x_3(v, \neg c_v) \geq 0 \quad (12)$$

$$\forall v : v\text{-np}, \quad \forall v\text{-np} \in I$$

$$x_3(v, c_v) - x_1(v\text{-np}, c) \geq 0 \quad (13)$$

$$\forall v : v\text{-np}, \quad \forall v\text{-np} \in I$$

Here $L_3 = \{c_v, \neg c_v\}$, $x_3(v, l)$ is the decision variable set to 1 only if the label $l \in L_3$ is assigned to v. The Equation 10 constraints that only one label out of $|L_3|$ choices can be assigned to a v. The equation 11 requires $x_3(v, l)$ to be a binary variable. The constraint 12 assumes if a verb v belongs to the class c_v (i.e., has least potential to encode causation) then its corresponding pair v-np encodes non-causality. The constraint 12 enforces that if a verb v belongs to the class $\neg c_v$ then its corresponding v-np pair is assigned the label $\neg c$. Similarly, the constraint 16 enforces that if a v-np pair encodes causality then its verb v has potential to encode causal relation. We maximize the objective function Z_3 subject to the constraints introduced above.

3.4 Knowledge of Indistinguishable Verb and Noun

As introduced earlier, each causal relation is characterized by two roles i.e., cause and its effect. In order to encode causal relation, two components of an instance of verb-noun_phrase pair need to represent distinct events, processes or phenomena. Employing simple lexical matching, we determine if a verb and a noun phrase represent same event or not as follows:

- We use NOMLEX (Macleod et al., 2009) to transform a verb into its corresponding nominalization and use the following text segments for lexical matching.

$$T_v = [\text{Subject}] \text{ verb } [\text{Object}]^2$$

$$T_n = \text{Head noun of noun phrase}$$

- We remove stopwords and duplicate words from T_v and T_n and take lemmas of all words. If the subject or object or both arguments are contained in noun phrase then we remove these arguments from T_v . We determine the probability of a verb (v) and a noun phrase (np) representing same event as follows. If head noun (i.e., T_n) lexically matches with any word of T_v then set $P(v \equiv np)$ to 1 and 0 otherwise.

We assign non-cause relation if $P(v \equiv np) = 1$. Next, we incorporate the knowledge of indistinguishable verb and noun in our model using the following additions to our integer linear program.

$$Z_4 = Z_3 + \sum_{v\text{-np} \in I} \sum_{l \in L_4} x_4(v\text{-np}, l) P(v\text{-np}, l) \quad (14)$$

$$\sum_{l \in L_4} x_4(v\text{-np}, l) = 1 \quad \forall v\text{-np} \in I \quad (15)$$

$$x_4(v\text{-np}, l) \in \{0, 1\} \quad \forall v\text{-np} \in I, \forall l \in L_4$$

$$x_1(v\text{-np}, -c) - x_4(v\text{-np}, \equiv) \geq 0 \quad \forall v\text{-np} \in I \quad (16)$$

Here $L_4 = \{\equiv, \neq\}$ where the label \equiv (\neq) represents same (distinct) events, $x_4(v\text{-np}, l)$ is the decision variable set to 1 only if the label $l \in L_4$ is assigned to v-np. The Equation 15 constraints that only one label out of $|L_4|$ choices can be assigned to a v-np pair. The equation 16 requires $x_4(v\text{-np}, l)$ to be a binary variable. The constraint 16 enforces that if a v-np pair belongs to the class \equiv then this pair is assigned the label $-c$. We maximize the objective function Z_4 subject to the constraints introduced above.

4 Evaluation and Discussion

In this section we present the experiments, evaluation procedures, and a discussion on the results achieved through our model for the current task.

In order to evaluate our model, we generated a test set with instances of form verb-noun_phrase where the verb is grammatically connected to the noun phrase in an instance. For this purpose, we

²Following Riaz and Girju (2010), we assume that the subject and object of a verb are parts of an event represented by a verb. Therefore, we use these arguments along with a verb for lexical matching with a noun phrase.

collected three wiki articles on the topics of Hurricane Katrina, Iraq War and Egyptian Revolution of 2011. We selected first 100 sentences from these articles and applied part-of-speech tagger (Toutanova et al., 2003) and dependency parser (Marneffe et al., 2006) on these sentences. Using each sentence, we extracted all verb-noun_phrase pairs where the verb has a dependency relation with any word of noun phrase. We manually inspected all of the extracted instances and removed those instances in which a word had been wrongly classified as a verb by the part-of-speech tagger. There are total 1106 instances in our test set. We assigned the task of annotation of these instances with cause and non-cause relations to a human annotator. Using manipulation theory of causality (Woodward, 2008), we adopted the annotation guidelines from Riaz and Girju (2010) which is as follows: “Assign cause label to a pair (**a**, **b**), if the following two conditions are satisfied: (1), **a** temporally precedes/overlap **b** in time, (2) while keeping as many state of affairs constant as possible, modifying **a** must entail predictably modifying **b**. Otherwise assign non-cause label.”

We have 149 (957) cause (non-cause) instances in our test set³, respectively. We evaluate the performance of our model using F-score and accuracy evaluation measures (see table 2 for results).

The results in table 2 reveal that the basic supervised classifier is a naive model and achieves only 27.27% F-score and 46.47% accuracy. The addition of novel types of knowledge introduced in section 3 (i.e., the model Basic+SCN_M+SCV+IVN) brings 14.74% (29.57%) improvements in F-score (accuracy), respectively. These results show that the knowledge of semantics of nouns and verbs and the knowledge of indistinguishable verb and noun are critical to achieve performance. The maximum improvement in results is achieved with the addition of semantic classes of nouns (i.e., Basic+SCN_M). The consideration of association of metonymic readings using model Basic+SCN_M helps us to maintain recall as compared with SCN_{!M} and therefore brings better F-score.

One can notice that almost all models suffer from low precision which leads to lower F-scores. Although, our model achieves 14.58% increase in precision over basic supervised classifier, the lack of high precision is still responsible for lower F-

³We will make the test set available

Model	Basic	+SCN _{!M}	+SCN _M	+SCN _M +SCV	+SCN _M +SCV+IVN
Accuracy	46.47	75.76	74.41	75.31	76.04
Precision	16.69	28.14	29.53	30.47	31.27
Recall	74.49	50.66	64.00	64.00	64.00
F-score	27.27	39.19	40.42	41.29	42.01

Table 2: This table presents results of the basic supervised classifier (i.e., Basic) and the models after incrementally adding the knowledge of semantic classes of nouns without consideration of metonymic readings (i.e., +SCN_{!M}), the knowledge of semantic classes of nouns with consideration of metonymic readings (i.e., +SCN_M), the knowledge of semantic classes of verbs (i.e., +SCN_M+SCV) and the knowledge of indistinguishable verb and noun (i.e., +SCN_M+SCV+IVN).

score. The highly skewed distribution of test set with only 13.47% causal instances results in lots of false positives. We manually examined false positives to determine the language features which may help us reducing more false positives without affecting F-score. We noticed that the direct objects of the verbs are mostly part of the event represented by the verbs and therefore encodes non-causation with the verbs. For example, consider following instances:

6. The hurricane surge protection failures prompted a lawsuit.
7. They provided weather forecasts.

In example 6, “lawsuit” is the direct object of the verb “prompted” and is part of the event represented by the verb “prompt”. However there is a cause relation between “protection failures” and “prompted”. Similarly in example 7, the direct object “forecasts” is part of the “providing” event and thus the noun phrase “weather forecasts” encode non-cause relation with the verb “provide”. Therefore, following this observation we employed the training corpus of cause and non-cause relations (see section 3.1) and learned the structure of verb-noun_phrase pairs encoding non-cause relations most of the time. We considered only those training instances where the subject and/or object of the verb was available. For the current purpose, we picked up following four features (1) sub_in_np, (2) !sub_in_np, (3) obj_in_np and (4) !obj_in_np. Just to remind the reader, the feature sub_in_np (!sub_in_np) is set to 1 if the subject of the verb is (not) contained in the noun phrase np, respectively. For each of the above four features, the percentage of cause and entropy of relations with that feature are as follows:

- sub_in_np (%c = 34.72, Entropy = 0.931)
- !sub_in_np (%c = 59.71, Entropy = 0.972)
- obj_in_np (%c = 28.89, Entropy = 0.867)

- !obj_in_np (%c = 55.30, Entropy = 0.991).

There are two important observations from above scores: (1) verbs mostly encode non-cause relations with their objects and subjects (i.e., high %-c with obj_in_np and sub_in_np), (2) among obj_in_np and sub_in_np features, obj_in_np yields least entropy i.e., there are least chances of encoding causality of a verb with its object.

Considering the above statistics, we enforce the constraint on each verb-noun_phrase pair that if the object of the verb is contained in the noun phrase of the above pair then assigns non-cause relation to that pair. Using this constraint, we obtain 46.61% (80.74%) F-score (accuracy), respectively. This confirms our observation that the object of a verb is normally part of an event represented by the verb and thus it encodes non-cause relation with the verb.

In this research, we have utilized novel types of knowledge to improve the performance of our model. In future, we need to consider more additional information (e.g., predictions from metonymy resolver) to achieve further progress.

5 Conclusion

In this paper, we have proposed a model for identifying causality in verb-noun pairs by employing the knowledge of semantic classes of nouns and verbs and the knowledge of indistinguishable noun and verb of an instance along with shallow linguistic features. Our empirical evaluation of model has revealed that such novel types of knowledge are critical to achieve a better performance on the current task. Following the encouraging results achieved by our model, we invite researchers to investigate more interesting types of knowledge in future to make further progress on the task of recognizing causality.

References

- Collin F. Baker, Charles J. Fillmore and John B. Lowe. 1998. The Berkeley FrameNet project. In *proceedings of the Association for Computational Linguistics and International Conference on Computational Linguistics (COLING-ACL)*.
- Brandon Beamer and Roxana Girju. 2009. Using a Bigram Event Model to Predict Causal Potential. In *proceedings of the Conference on Computational Linguistics and intelligent Text Processing (CICLING)*.
- Steven Bethard and James H. Martin. 2006. Identification of Event Mentions and their Semantic Class. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- Steven Bethard and James H. Martin. 2008. Learning Semantic Links from a Corpus of Parallel Temporal and Causal Relations. In *proceedings of the Association for Computational Linguistics (ACL)*.
- Du-Seong Chang and Key-Sun Choi. 2006. Incremental cue phrase learning and bootstrapping method for causality extraction using cue phrase and word pair probabilities. *Information Processing and Management, volume 42 issue 3, 662678*.
- Timothy Chklovski and Patrick Pantel. 2004. VerbOcean: Mining the Web for Fine-Grained Semantic Verb Relations. In *proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- Quang X. Do, Yee S. Chen and Dan Roth. 2011. Minimally Supervised Event Causality Identification. In *proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- Jenny R. Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. In *Proceedings of the Association for Computational Linguistics (ACL)*.
- Roxana Girju. 2003. Automatic detection of causal relations for Question Answering. *Association for Computational Linguistics ACL, Workshop on Multilingual Summarization and Question Answering Machine Learning and Beyond*.
- Roxana Girju and Dan Moldovan. 2002. Mining Answers for Causation Questions. In *American Associations of Artificial Intelligence (AAAI), 2002 Symposium*.
- Jaegwon Kim. 1993. Causes and Events. Mackie on Causation. In *Causation, Oxford Readings in Philosophy, ed. Ernest Sosa, and Michael Tooley, Oxford University Press*.
- Catherine Macleod, Ralph Grishman, Adam Meyers, Leslie Barrett, Ruth Reeves. 1998. NOMLEX: A Lexicon of Nominalizations. In *proceedings of EURALEX*.
- Katja Markert, Malvina Nissim. 2009. Data and models for metonymy resolution. *Language Resources and Evaluation Volume 43 Issue 2, Pages 123–138*.
- Marie-Catherine de Marneffe, Bill MacCartney and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*.
- Emily Pitler, Annie Louis and Ani Nenkova. 2009. Automatic Sense Prediction for Implicit Discourse Relations in Text. In *proceedings of ACL-IJCNLP*.
- Emily Pitler and Ani Nenkova. 2009. Using Syntax to Disambiguate Explicit Discourse Connectives in Text. In *proceedings of ACL-IJCNLP*.
- James Pustejovsky, Patrick Hanks, Roser Saur, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro and Marcia Lazo. 2003. The TIMEBANK Corpus. In *Proceedings of Corpus Linguistics*.
- Kira Radinsky and Eric Horvitz. 2013. Mining the Web to Predict Future Events. In *proceedings of sixth ACM international conference on Web search and data mining, (WSDM)*.
- Mehwish Riaz and Roxana Girju. 2010. Another Look at Causality: Discovering Scenario-Specific Contingency Relationships with No Supervision. In *proceedings of the IEEE 4th International Conference on Semantic Computing (ICSC)*.
- Mehwish Riaz and Roxana Girju. 2013. Toward a Better Understanding of Causality between Verbal Events: Extraction and Analysis of the Causal Power of Verb-Verb Associations. *Proceedings of the annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL)*.
- Dan Roth and Wen-tau Yih. 2004. A Linear Programming Formulation for Global Inference in Natural Language Tasks. In *Proceedings of the Annual Conference on Computational Natural Language Learning (CoNLL)*.
- Caroline Sporleder and Alex Lascarides. 2008. Using automatically labelled examples to classify rhetorical relations: An assessment. *Journal of Natural Language Engineering Volume 14 Issue 3*.
- Kristina Toutanova, Dan Klein, Christopher Manning, and Yoram Singer. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In *Proceedings of Human Language Technology and North American Chapter of the Association for Computational Linguistics (HLT-NAACL)*.
- James Woodward. 2008. Causation and Manipulation. *Online Encyclopedia of Philosophy*.

Appendix A. Semantic Classes of Nouns

This appendix presents the FrameNet labels we assign to c_{np} and $\neg c_{np}$ classes (see section 3.2).

Semantic Class	FrameNet Labels
c_{np}	Event, Goal, Purpose, Cause, Internal cause, External cause, Result, Means, Reason, Phenomena, Characterization, Coordinated event, Final state, Information, Topic, Containing event, Mental content, Action, Experience, Impactee, Impactor, Message, Question, Circumstances, Desired goal, Explanation, Required situation, Complaint, Content, Activity, Intended goal, Phenomenon, State, Dependent state, Forgery, Purpose of Event, Negative consequences, Inference, Appraisal, Noisy event, Function, Evidence, Process, Paradigm, Standard, Old order, Focal occasion, Landmark occasion, resulting action, Victim, Issue, Effect, State of affairs, Cause of shine, Qualification, Undesirable Event, Skill, Precept, Outcome, Norm, Act, State of Affairs, Phenomenon 1, Phenomenon 2, Quality Eventuality, Expression, Intended event, Cognate event, Epistemic stance, Goal conditions, Possession, Support Proposition, Domain of Relevance, Charges, Idea, Initial subevent, Hypothetical event, Scene, Purpose of Goods, Response action, Motivation, Executed, Affliction, Medication, Treatment, Stimulus, Last subevent, Undesirable situation, Sleep state, Initial state, Enabled situation, Grinding cause, Finding, Case, Legal Basis, Role of focal participant, Trigger, Authenticity, World state, Emotion, Emotional state, Evaluation, New idea, Production, Performance, Undertaking, Destination event
$\neg c_{np}$	Artist, Performer, Duration, Time, Place, Distributor, Area, Path, Direction, Sub-region, Creator, Copy, Original, Iteration, Manner, Frequency, Agent, Body part, Depictive, Theme, Subregion, Area, Degree, Angle, Fixed location, Path shape, Addressee, Entity, Individual 1, Individual 2, Road, Distance, Speaker, Medium, Clothing, Wearer, Bodypart of agent, Locus, Cognizer, Salient entity, Name, Inspector, Ground, Unwanted entity, Location of inspector, Researcher, Population, Searcher, Sought entity, Instrument, Created entity, Components, Forgoer, Desirable, Bad entity, Dodger, Experiencer, Vehicle, Self mover, Speed, Cotheme, Consecutive, Re encoding, Supplier, Individuals, Driver, Complainer, Communicator, Protagonist, Attribute, Final value, Item, Initial value, Difference, Group, Value range, Co participant, Perceiver agentive, Target symbol, Location of perceiver, Location, Expected entity, Focal participant, Time of Event, Variable, Limits, Limit1, Limit2, Point of contact, Goods, Lessee, Lessor, Money, Rate, Unit, Reversible, Perceiver passive, Sound, Sound source, Location of source, Fidelity, Official, Selector, Role, Concessive, New leader, Body, Old leader, Leader, Governed, Result size, Size change, Dimension, Initial size, Elapsed time, Interval, Category, Criteria, Text, Final correlate, Correlate, Initial correlate, Manipulator, Side 1, Sides, Side 2, Perpetrator, Value 1, Value 2, Actor, Partner 2, Partner 1, Partners, Figure, Resident, Co resident, Student, Subject, Institution, Level, Teacher, Undergoer, Subregion bodypart, Course, Owner, Defendant, Judge, Co abductee, Location of appearance, Material, Accused, Arraign authority, Hair, Configuration, Emitter, Beam, Amount of progress, Evaluatee, Patient, Buyer, Seller, Recipient, Relay, Relative location, Connector, Items, Part 1, Part 2, Parts, Whole, Name source, Payer, Fine, Executioner, Interlocutor 1, Interlocutor 2, Interlocutors, Healer, Food, Cook, Container, Heating instrument, Temperature setting, Resource controller, Resource, Donor, Constant location, Carrier, Sender, Co theme, Transport means, Holding location, Rope, Knot, Handle, Containing object, Fastener, Enclosed region, Container portal, Aggregate, Suspect, Authorities, Offense, Source of legal authority, Ingestor, Ingestibles, Sleeper, Pieces, Goal area, Period of iterations, Mode of transportation, Produced food, Ingredients, Cognizer agent, Excreter, Excreta, Air, Perceptual source, Escapee, Undesirable location, Evader, Capture, Pursuer, Amount of discussion, Means of communication, Periodicity, Author, Honoree, Reader, Child, Mother, Father, Egg, Flammables, Flame, Kindler, Mass theme, Address, Intermediary, Communication, Location of communicator, Firearm, Indicated entity, Hearer, Sub region, Member, Object, Organization, Guardian, New Status, Arguer, Criterion, Liquid, Impactors, Force, Coparticipant, Holding Location, Legal basis, Precipitation, Quantity, Voice, Duration of endstate, Period of Iterations, Employer, Employee, Task, Position, Compensation, Field, Place of employment, Amount of work, Contract basis, Recipients, Hot Cold source, Temperature goal, Temperature change, Hot/Cold source, Dryee, Temperature, Traveler, Iterations, Baggage, Deformer, Resistant surface, Fluid, Injured Party, Avenger, Injury, Punishment, Offender, Grinder, Profiled item, Standard item, Profiled attribute, Standard attribute, Extent, Source emitter, Emission, Sub source, Item 1, Item 2, Parameter, Form, Chosen, Change agent, Injuring entity, Severity, Substance, Delivery device, Entry path, Wrong, Amends, Grounds, Expressor, Basis, Signs, Manufacturer, Product, Factory, Consumer, Interested party, Performer1, Performer2, Whole patient, Destroyer, Exporting area, Importing area, Accuracy, Time of Eventuality, Indicator, Indicated, Audience, Valued entity, Journey, Duration of end state, Killer, Beneficiary, Destination time, Landmark time, Seat of emotion, Arguers, Arguer1, Arguer2, Company, Asset, Origin, Sound maker, Static object, Themes, Heat source, Following distance, Perceiver, Intended perceiver, Location of expressor, Path of gaze, Relatives, Final temperature, Particular iteration, Participant 1, Language

Table 3: This table presents the FrameNet labels we assign to c_{np} and $\neg c_{np}$ classes.