Semantic Novelty Detection and Characterization in Factual Text Involving Named Entities

Nianzu Ma[†], **Sahisnu Mazumder**[‡], **Alexander Politowicz**[†],
Bing Liu[†], **Eric Robertson**[♮], **Scott Grigsby**[♮]
[†]Department of Computer Science, University of Illinois at Chicago, USA
[‡]Intel Labs, USA
[♮]PAR Government Systems Corporation, USA
jingyima005@gmail.com, sahisnumazumder@gmail.com, politow2@uic.edu
liub@uic.edu, {eric_robertson,Scott_Grigsby}@partech.com

Abstract

Much of the existing work on text novelty detection has been studied at the topic level, i.e., identifying whether the topic of a document or a sentence is novel or not. Little work has been done at the fine-grained *semantic level* (or *contextual level*). For example, given that we know Elon Musk is the CEO of a technology company, the sentence "*Elon Musk acted in the sitcom The Big Bang Theory*" is novel and surprising because normally a CEO would not be an actor. Existing topic-based novelty detection methods work poorly on this problem because they do not perform semantic reasoning involving relations between named entities in the text and their background knowledge. This paper proposes an effective model (called PAT-SND) to solve the problem, which can also characterize the novelty. An annotated dataset is also created. Evaluation shows that PAT-SND outperforms 10 baselines by large margins.

1 Introduction

Novelty/anomaly detection has been an active research area for decades (Grubbs, 1969; Chalapathy et al., 2018; Pang et al., 2021). Recently, it has received increased attention in NLP. Broadly speaking, there are two main types: **(1) Topic-based novelty detection**, which classifies a given text to a training/known class (topic) or reject/detect it as belonging to some unknown classes (Fei and Liu, 2016; Shu et al., 2017; Lin and Xu, 2019; Zheng et al., 2020); and **(2) Semantic novelty detection**, which determines whether a given text represents a semantically/contextually novel phenomenon. For example, Ma et al. (2021) studied a semantic novelty detection problem - detecting semantically novel scene descriptions (e.g., "*A person walks a chicken in the park*" is a novel scene, whereas "*A person walks a dog in the park*" is normal one). This task is more fine-grained and requires factual reasoning over text as compared to that of (1), which has been studied extensively. (2)

d_1 :	"*The Big Bang Theory is an American television sitcom, filmed in front of a live audience, stars Johnny Galecki et al.*"	**normal**
d_2 :	"*Elon Musk's performance as a dishwasher in a restaurant in season 9, episode 9 of the The Big Bang Theory is quite interesting to his fans.*"	**novel**

Figure 1: Examples of semantic novelty detection in factual texts involving named entities.

has only been introduced recently and is the focus of this paper.

This paper proposes a new semantic novelty detection task: given a factual text d containing two *named entities*[1], we want to classify whether d represents a semantically novel fact or a normal one *with respect to* the entity pair. For example, consider the text d_1 and an entity pair underlined in d_1 in Figure 1. d_1 represents a *normal* fact as it is natural for an actor (*Johnny Galecki*) to act in a sitcom or TV show (*The Big Bang Theory*). However, d_2 in Figure 1 depicts a novel fact with respect to the underlined entity pair because a CEO of a technology company (*Elon Musk*) acting in a sitcom (*The Big Bang Theory*) is quite surprising and novel.

Factual text appears in diverse media sources, such as news articles, blog posts, reviews etc. Detecting semantically novel facts involving popular real-world (named) entities has many applications because anything novel is always of interest and can trigger readers' curiosity. For example, a mobile newsfeed application can increase user engagement by recommending novel news/facts of named entities and promoting news articles with novel facts. Although novelty is subjective and personal, there exist some novel facts that the majority of people agree. In this work, we restrict our study to this consensus-view of semantic novelty and leave the personalized novelty for future work.

[1]Named Entity definition: https://en.wikipedia.org/wiki/Named_entity

Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pages 9225 - 9252
December 7-11, 2022

Solving the proposed task requires *joint* fine-grained reasoning over (1) the *relationship* between the pair of entities in the textual context and (2) the *background knowledge* of the entities. For example, considering d_2 in Figure 1, we first need to detect that the entity pair ("*Elon Musk*", "*The Big Bang Theory*") in d_2 has the "*cast-member*" relation and then, leverage the interaction of the relation with the background knowledge of the entities (i.e., "*Elon Musk*" is a *tech entrepreneur* and "*The Big Bang Theory*" is a *TV show*) to infer the semantic novelty (because, a tech entrepreneur does not normally act in a TV show). We utilize the external Knowledge Repository (KR) - WikiData (Vrandečić and Krötzsch, 2014) to extract the named entity's background knowledge, which is a list of property-value pairs. For example, *Elon Musk*'s background knowledge contains property-value pairs: [(a) (`occupation`, *entrepreneur*), (b) (`gender`, *male*), (c) (`field of work`, *tech entrepreneurship*) ....]. However, not all property-value pairs is useful for inference (e.g., only (a) and (c) are useful for d_2 in Figure 1). Thus, a solution for automatic selection of the *useful* property-value pairs is needed (see Sec. 4). In fact, the *useful* property-value pairs provide a reason or *characterization* for the novelty.

Problem Definition: Given (1) a set of training factual text $\mathcal{D}_{tr} = \{d_1, d_2, ...d_n\}$, with each $d_i \in \mathcal{D}_{tr}$ labeled as normal (*NORMAL* class) with respect to a pair of entities (e_1^i, e_2^i) appeared in d_i, and (2) a knowledge base (KB) $\mathcal{K}$ containing the background knowledge (property-value pairs) of a set of entities that is a superset of the entities appeared in $\mathcal{D}_{tr}$, our goal is to build a model $\mathcal{F}$ to score the semantic novelty of a test factual text d' having a pair of entities ($e_1^{'}$, $e_2^{'}$) with respect to $\mathcal{D}_{tr}$, $\mathcal{K}$, and pair ($e_1^{'}$, $e_2^{'}$), i.e., classifying d' into one of the classes {*NORMAL*, *NOVEL*}. As $\mathcal{F}$ is built with only the "*NORMAL*" data, the task is an *one-class classification problem.*

This task is different from the semantic novelty detection task in (Ma et al., 2021) in two main aspects: (1) Our task demands semantic reasoning over named entities which do not have sufficient semantic information in their textual (or surface) form in d. Rich background knowledge of the entities is needed to detect novelty. The task in (Ma et al., 2021) does not require any of such entity background knowledge. (2) (Ma et al., 2021) do semantic reasoning for relations (between objects), based on a fixed/closed set of verbs. However, in our work, the relations between entities may be expressed in any surface forms and/or even implicitly (e.g., the relation "*cast-member*" between the underlined entities is expressed implicitly in d_2). Ma et al. (2021) cannot handle such cases.

To solve the task, we propose a new model, called PAT-SND (***P**roperty **AT**tention network for **S**emantic **N**ovelty **D**etection*) to detect novel factual text. Additionally, PAT-SND also provides the characterization (or reason) for the novelty (unlike Ma et al. (2021)). PAT-SND first employs an existing relation classification technique to identify the relation between the entity pair. The identified relation is then used in a novel *relation-aware* **P**roperty **AT**tention Network (PAT) module that leverages the attention mechanism to select the useful background knowledge from the KB $\mathcal{K}$ to perform semantic reasoning for novelty detection. The learned attention knowledge in PAT is also used to provide the characterization for the novelty (see Sec. 4).

PAT-SND is evaluated using our **newly created** NFTD (**N**ovel **F**actual **T**ext **D**etection) dataset. We leverage a distant supervision technique (Mintz et al., 2009) with the Wikipedia[2] as the corpus and Wikidata as the KR to build a large training dataset. Evaluation results show that PAT-SND outperforms the 10 latest novelty detection baselines by very large margins.

Our main contributions are as follows:

1. We propose a new semantic novelty detection task for factual text involving named entities.

2. We propose an effective technique called PAT-SND to solve the proposed task.

3. The proposed technique also provides the characterization of novelty based on the attention knowledge in the PAT-SND model.

4. A new dataset called NFTD is created for the proposed task as no suitable data is available. The dataset can be used as a benchmark by the NLP community.

2 Related Work

Novelty or anomaly detection has been studied extensively over the years. Early representative works include *one-class SVM* (OCSVM) (Schölkopf et al.,

[2] https://en.wikipedia.org/wiki/Main_Page

2001; Manevitz and Yousef, 2001), *Support Vector Data Description* (SVDD) (Tax and Duin, 2004) and hybrid approaches (Erfani et al., 2016; Ruff et al., 2018) that learn features using deep learning and then apply OCSVM or SVDD to build one-class classifiers. More recent deep learning approaches are based on auto-encoders (You et al., 2017; Abati et al., 2019; Chalapathy and Chawla, 2019), GAN (Perera et al., 2019; Zheng et al., 2019), neural density estimation (Wang et al., 2019), multiple hypothesis prediction (Nguyen et al., 2019), robust mean estimation (Dong et al., 2019) and regularization (Hu et al., 2020). Chalapathy and Chawla (2019); Pang et al. (2021) provides a detailed survey. Our PAT-SND is based on an attention network and data augmentation technique.

Novelty detection has also been studied in out-of-distribution (OOD) detection or open-set recognition (Liang et al., 2018; Shu et al., 2018; Erfani et al., 2017; Xu et al., 2019). However, these methods work in the multi-class setting. Ours is an one-class classification problem. There are also works on topical novelty detection (Dasgupta and Dey, 2016; Ghosal et al., 2018; Nandi and Basak, 2020; Jo et al., 2020; Li and Croft, 2005; Zhang and Tsai, 2009). They differ from ours as we focus on *fine-grained* semantic novelty detection.

Our work is also related to *Semantic plausibility* (SPL) that studies the problem of whether an event is plausible or not (Porada et al., 2019; Wang et al., 2018; Keller and Lapata, 2003; Zhang et al., 2017; Sap et al., 2019) and *selectional preference* (SPR) that deals with the "typicality" of an event (Resnik, 1996; Clark and Weir, 2001; Erk and Padó, 2010; Bergsma et al., 2008; Ritter et al., 2010; Ó Séaghdha, 2010; Van de Cruys, 2009, 2014; Dasigi and Hovy, 2014; Tilk et al., 2016). These works differ from ours as (1) conceptually, SPL and SPR are related but different from novelty, (2) Our task demands the use of background knowledge in the named entities for semantic reasoning. However SPL and SPR only perform reasoning on the surface form of objects in the text, and (3) they use fully labeled data (Dasigi and Hovy, 2014) while we have only normal (one-class) data in training.

Commonsense reasoning is remotely related to our work. Existing works have built multi-choice commonsense reasoners (Zellers et al., 2018, 2019), studied the commonsense knowledge contained in language models (Davison et al., 2019; Trinh and Le, 2019, 2018) and knowledge graph (Bosselut et al., 2019), and constructed new datasets for better evaluation (Wang et al., 2020a). Several researchers also investigated physical commonsense reasoning (Bagherinezhad et al., 2016; Forbes and Choi, 2017; Wang et al., 2017; Bisk et al., 2020) and affordance of entities (Forbes et al., 2019). They do not perform novelty detection.

Trivia fact mining (Merzbacher, 2002; Ganguly et al., 2014; Gamon et al., 2014; Prakash et al., 2015; Fatma et al., 2017; Mahesh and Karanth; Tsurel et al., 2017; Niina and Shimada, 2018; Korn et al., 2019; Kwon et al., 2020) is also relevant, but it is mainly about interestingness. Some trivia facts are interesting because they are rare, *but not necessarily novel*. Existing papers use labeled training data for learning, or rely on Wikipedia structure to retrieve interesting facts using information retrieval methods (Tsurel et al., 2017; Kwon et al., 2020). We have only normal data but no novel data.

Our proposed model is based on an attention network. Related NLP works using attention techniques include (Huang and Carley, 2019; Ma et al., 2020; Guo et al., 2019; Wang et al., 2020b; Pouran Ben Veyseh et al., 2020; Xiao and Zhou, 2020). But they solve different problems, such as sentiment analysis and argument mining and are not about novelty detection. Their approaches also differ from ours.

3 Dataset Collection and Annotation

To build a large factual text dataset annotated with named entities, we leverage the distant supervision technique in Mintz et al. (2009). We create our training and test datasets, using Wikipedia as the corpus and Wikidata (Vrandečić and Krötzsch, 2014) as the external Knowledge Repository (KR).

We choose Wikidata as KR for extracting background information of the entities, because the good community collaboration and contribution of Wikidata makes it a high-quality KR compared to other KRs (Färber et al., 2015). Wikidata encodes real-world knowledge in the form of triples: (e_1, r, e_2), which means entity e_1 and entity e_2 have a relation r. For instance, (The Big Bang Theory, Cast-Member, Johnny Galecki).

The named entities in the Wikipedia corpus are linked to the Wikidata. We can find unambiguous mappings between entity mentions in the text and Wikidata entities. For example: In the *Wikipedia Source*: "[[The Big Bang Theory]] is an American television sitcom, filmed in front of a live audience,

Table 1: NFTD dataset statistics. NR (NV) denotes the NORMAL (NOVEL) class. "text length" is # of words.

	Training	Test
# instances (factual text)	251,619 (NR)	1000 (NR), 1000 (NV)
Avg. text length	41.35	26.02

stars [[Johnny Galecki]] et al.", the named entities in bracket [[.]] have an unique one-to-one mapping to the entities in Wikidata.

Training dataset preparation. The distant supervision technique can be briefly described as follows: For a piece of text d from Wikipedia involving e_1 and e_2 (with hyperlink uniquely mapping to Wikidata entities), if there is a triple (e_1, r, e_2) in the KR, we assume that the textual information in d expresses the relation r between e_1 and e_2. In this case, we automatically annotate (e_1, r, e_2, d) as a distantly supervised instance and add it to our training dataset. For entity pairs (e_1, e_2) with more than one relation, we discard them because they bring ambiguity in our dataset.

Due to the budgetary constraints, we can not evaluate on all relations in the Wikidata. We create our training data related to 20 human related relations. The details of these 20 relations are in Appendix Sec. A. With distant supervision, we allow noise to exist in the training dataset because this process requires no human annotation, and scales up the learning of more relations. We split the whole dataset created via distant supervision into two parts: *train set* and the *test set pool*, making sure that there is no overlapping in either text or entity pairs between these two parts. This *test set pool* is used for test dataset preparation.

Test dataset preparation. While training dataset may contain noise, test data needs to be manually annotated and checked for a fair evaluation. We invited five graduate students with advanced level of English as crowd workers. We randomly split the *test set pool* into two parts: *normal test data pool-1* and *normal test data pool-2*.

Normal test data. We assume that the fact descriptions in Wikipedia are all normal facts. So for normal test data, we sample instances from the *normal test data pool-1* and assign them to annotators to identify the correct instances. Each instance is a tuple (e_1, r, e_2, d). The annotators are asked to check whether or not the sentence d with the entity pair (e_1, e_2) semantically expresses the relation r. If yes, this instance is added to our normal test dataset. After an instance is collected, we ensure that it is verified by another annotator. If there is a disagreement, we make sure it is discussed and resolved between the two annotators. Following this procedure, we annotate 50 normal instances for each relation.

Novel test data. We divide the whole task into 20 subtasks and evenly assign them to the annotators. For each subtask, the goal is to generate 50 novel tuples (e_1, r, e_2, d) for each relation. Instead of asking annotators to create novel instances from scratch, we sample some instances from *normal test data pool-2* to inspire annotators. They are asked to change the property-value pairs of entities and the text d, or even write from scratch if they come up with interesting ideas.

After the first round annotation, we get 50 novel instances for each of the 20 relations. Then, the annotations are shown to the other four annotators to label them as normal or novel. We use the majority voted label as the final label of these instances. We use Fleiss' Kappa (Fleiss and Cohen, 1973) to calculate the inter-rater reliability. The Fleiss' Kappa score is 0.91, interpreted as high agreement, which means our test data reflects the consensus-view of semantic novelty. At the end, we collect 50 normal and 50 novel instances for each of the 20 relations. Table 1 shows the NFTD dataset statistics.

The details of the data annotation guideline is in Appendix Sec. C.

Building Entity Background KB ($\mathcal{K}$). We use the knowledge repository (KR), Wikidata, to build the entity background KB $\mathcal{K}$. KR is represented as: $KR = (\mathcal{E}, \mathcal{R}, \mathcal{T})$, where $\mathcal{E}$ denotes a set of entities, $\mathcal{R}$ is a set of relations/edges, and $\mathcal{T} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ is the set of all triples. For each entity e in $\mathcal{E}$, we obtain the list of property-value pairs as e's background knowledge to build $\mathcal{K}$ as follows.

We first collect all triples from KR involving e and then extract the relation and the other entity from each triple to form a property-value pair with the relation as a property and the other entity as the value of the property. For example, considering e = "*Elon Mask*" and a triple ("*Elon Mask*", "*occupation*", "*entrepreneur*") in KR, the extracted property-value pair for e would be (`occupation`, *entrepreneur*).

Let $\mathcal{P}$ be the complete property set in the background KB $\mathcal{K}$. We assume that each e_i in the training data is in the $\mathcal{K}$. However, e_i in the test data can be a new entity (i.e., it does not appear in the training data), as long as the background knowledge of the entity is available to our model (where, the property-value pairs are either retrieved from the KR or provided by the human annotator during the test data annotation process and included in $\mathcal{K}$).

4 Proposed Approach

Our proposed PAT-SND model works in two steps: (1) *Entity Relation Classification*, and (2) *Triple Semantic Novelty Scoring* (SNS). Given a factual text d containing a pair of entities (e_1, e_2), PAT-SND first identifies the relation $\hat{r}$ between (e_1, e_2) in d in step (1) [Sec. 4.1]. Next, the background knowledge of the entities e_1 and e_2 retrieved from the KB $\mathcal{K}$ together with the predicted relation $\hat{r}$ are fed to the SNS module to score the semantic novelty of d with respect to (e_1, e_2) and $\mathcal{K}$ in step 2 [Sec. 4.2]. As our training data $\mathcal{D}_{tr}$ consists of only NORMAL class examples (as discussed in Sec. 1), it's not possible to train SNS solely with $\mathcal{D}_{tr}$. Thus, we propose a *KB-based Contrastive Data Generator* (CDG) to generate pseudo-novel examples. The SNS module is then trained with both NORMAL class examples in $\mathcal{D}_{tr}$ as well as the generated pseudo-novel examples in a supervised learning manner. We will discuss more about it in Sec. 4.3.

4.1 Entity Pair Relation Classification

Given a factual text d having entity pair (e_1, e_2), we build a model to identify the relation $\hat{r}$ between $(e_1$ and $e_2)$ in d. For this purpose, we utilize a BERT-based Relation Classification model (Wu and He, 2019), that incorporates entity position information into a pre-trained language model for relation classification. Next, we combine the identified relation $\hat{r}$ with the entity pair to produce a triple $(e_1, \hat{r}, e_2)$ which serves as input to the SNS (in Sec. 4.2).

During training process, the relation classification model is trained using $\mathcal{D}_{tr}$, where each $d_i \in \mathcal{D}_{tr}$ is labelled with *true relation label* r between the entity pair through the distant supervision technique.

4.2 Triple Semantic Novelty Scoring (SNS)

Let $B_1 = \{(p_i^1, v_i^1) | 1 \le i \le l\}$ and $B_2 = \{(p_i^2, v_i^2) | 1 \le i \le m\}$ be the background knowledge obtained for e_1 and e_2 respectively from KB $\mathcal{K}$ (See Sec. 3). The SNS module utilizes B_1, B_2 and relation $\hat{r}$ as inputs to score the novelty of the input text d. In this process, SNS employs a relation-aware attention mechanism over B_1 and B_2 to select the useful knowledge, which is motivated as follows.

e_1: The Big Bang Theory	
Property	*Value*
instance of	television series
start time	24 September 2007
end time	16 May 2019
audio system	Dolby Digital
…	

e_2: Elon Musk	
Property	*Value*
instance of	human
gender	male
occupation	entrepreneur
field of work	tech entrepreneurship
…	

Figure 2: Illustration of two entities' property and value pairs in the KB $\mathcal{K}$. The properties marked in red are useful or important for detecting semantic novelty of the example d_2 in Figure 1.

Leveraging all property-value pairs in B_1 and B_2 may not be helpful to detect the novelty of the text d. For example, as shown in Figure 2, considering the entity "*Elon Mask*", the property-value pair (`occupation`, *entrepreneur*) is useful to score the novelty of d_2 in Figure 1, whereas (`gender`, *male*) is not useful at all. Thus, the model needs to have the ability to focus on important information and filter out noises in B_1 and B_2. Such knowledge selection process is relation dependent, as for different relations, different property-value pairs would be useful for novelty detection.

To enable automated knowledge selection, SNS is built using a key component called Property Attention Network (PAT) that utilizes the semantics of the relation $\hat{r}$ to attend over B_1 and B_2 for inference. As the attention mechanism needs to be relation-specific, we build one PAT module for each relation. So, for detecting novelty of a test text d', SNS fires the PAT learned for relation $\hat{r}$, identified from d' using the Relation Classifier (in Sec. 4.1).

Property Attention Network (PAT). PAT takes a list of property-value pairs $\{(p_i, v_i) | 1 \le i \le N\}$ and a relation r as input and outputs a weighted value vector $\boldsymbol{h}^{out}$ to be used for inference. p_i and the corresponding v_i are fed to PAT as feature vectors $\boldsymbol{p}_i$, $\boldsymbol{v}_i$ respectively, together with r (to invoke the relation-specific module). We employ BERT (Devlin et al., 2019) to learn the embedding representation of p_i , v_i and use them as corresponding feature vectors. For example, the property "*instance of*" is encoded as $\langle$[CLS], instance, of, [SEP]$\rangle$ using WordPiece Tokenizer and fed into BERT and embedding corresponding to token [CLS] in the output layer of BERT is used as the feature vector of the property.

In PAT, the $\{\boldsymbol{p}_i\}_{i=1}^{N}$ are fed one by one through a relation-specific linear layer, and a *relu* non-linearity function and a softmax function are used to obtain the attention weights $\{\boldsymbol{\alpha}_{ir}\}_{i=1}^{N}$ over $\{\boldsymbol{p}_i\}_{i=1}^{N}$ with respect to r. Next, the weights are used to weigh the corresponding $\{\boldsymbol{v}_i\}_{i=1}^{N}$ to obtain $\boldsymbol{h}^{out}$. The processing for a given r is summarized:

$$\begin{aligned} \boldsymbol{g}_{ir}^{k} &= relu(\boldsymbol{p}_i\, \boldsymbol{W}_r^k + \boldsymbol{b}_r^k) \\ \boldsymbol{\alpha}_{ir}^{k} &= \frac{exp(\boldsymbol{g}_{ir}^{k})}{\sum_{i=1}^{N} exp(\boldsymbol{g}_{ir}^{k})} \\ \boldsymbol{h}^{out} &= \sum_{i=1}^{N}(\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{\alpha}_{ir}^{k})\, \boldsymbol{v}_i^k \end{aligned} \tag{1}$$

where K is the total number of attention heads and $\boldsymbol{W}_r^k$, $\boldsymbol{b}_r^k$ are relation-specific weight and bias for the k-th attention head. $\boldsymbol{\alpha}_{ir}^{k}$ is the k-th attention weight between r and p_i. Overall, the processing of inputs in PAT is denoted as $h^{out} = PAT(\boldsymbol{P}, \boldsymbol{V}, r; \Theta_r)$, where $\boldsymbol{P} = [\boldsymbol{p}_1, \boldsymbol{p}_2, ..., \boldsymbol{p}_N] \in \mathbb{R}^{N \times F}$ is the property matrix, $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_N] \in \mathbb{R}^{N \times F}$, is the value matrix and Θ_r is the trainable parameters for relation r.

Triple Novelty scoring. Given the inputs B_1, B_2 and relation $\hat{r}$, we obtain the property and value matrices $\boldsymbol{P_1}$, $\boldsymbol{V_1}$ from B_1 and $\boldsymbol{P_2}$, $\boldsymbol{V_2}$ from B_2 and feed them to PAT for relation $\hat{r}$ as follows:

$$\begin{aligned} h_1^{out} &= PAT(\boldsymbol{P_1}, \boldsymbol{V_1}, \hat{r}; \Theta_{\hat{r}}) \\ h_2^{out} &= PAT(\boldsymbol{P_2}, \boldsymbol{V_2}, \hat{r}; \Theta_{\hat{r}}) \\ h_{\hat{\tau}}^{out} &= [h_1^{out}; h_2^{out}] \end{aligned} \tag{2}$$

Next, a relation-specific feed-forward layer is used to project $h_{\hat{\tau}}^{out}$ into a semantic novelty score as $S(\hat{\tau}) = (h_{\hat{\tau}}^{out}\, \boldsymbol{W}_{\hat{r}} + \boldsymbol{b}_{\hat{r}})$, where $\hat{\tau}$ denotes the triple (e_1, $\hat{r}$, e_2). Following the existing one-class classification literature (Chalapathy and Chawla, 2019; Pang et al., 2021), we do not use a threshold to further produce a classification label, instead use $S(\hat{\tau})$ directly in our experiments (Sec. 5).

4.3 Training

Let $\mathcal{T}_{tr}$ be the set of all triples (labelled as NORMAL class) extracted from the examples in $\mathcal{D}_{tr}$. To train SNS, we use KB $\mathcal{K}$ to help generate contrastive examples (triples) by corrupting the triples in $\mathcal{T}_{tr}$, as discussed below. These contrastive examples serve as the pseudo-novel data and enable the supervised learning of the SNS.

KB-based Contrastive Data Generator. Given a triple $\tau_i \in \mathcal{T}_{tr}$, the generator $G_{contrastive}(\tau_i)$ randomly samples an entity $e^{'}$ from KB $\mathcal{K}$ to replace either e_1 or e_2 in τ_i. After corruption, $\tau_i^{'}$ is formed from τ_i, where $\tau_i^{'} = (e^{'}, r, e_2)$ or $\tau_i^{'} = (e_1, r, e^{'})$. For example, given τ_1 = (*The Big Bang Theory*, *cast-member*, *Johnny Galecki*) as a NORMAL triple in $\mathcal{T}_{tr}$, a pseudo-novel triple generated by $G_{contrastive}(\tau_1)$ would be $\tau_1^{'}$ = (*The Big Bang Theory*, *cast-member*, *Warren Buffett*). During the training of SNS, we dynamically generate one pseudo-novel triple for each NORMAL triple in $\mathcal{T}_{tr}$ in every training epoch.

Learning. PAT-SND is trained end-to-end by minimizing a max-margin ranking objective as,

$$\mathcal{L} = \sum_{\tau \in \mathcal{T}_{tr}} \sum_{\tau' \in \mathcal{T}_{tr}'} max\{S(\tau') - S(\tau) + 1, 0\} \tag{3}$$

where, $\mathcal{T}_{tr}'$ is the set of pseudo-novel triples generated from $\mathcal{T}_{tr}$. $\mathcal{L}$ encourages the score $S(\tau)$ of the NORMAL triple τ to be higher than $S(\tau')$ of a pseudo-novel triple τ'.

5 Experiments

5.1 Experiment Setup

The details of the dataset annotation and statistics have been discussed in Sec. 3. All the results reported in this section are the averages of five runs with different random seeds. The code and the dataset are released[3].

Evaluation Metrics. Since our task is an one-class classification task, we follow the existing one-class classification literature (Chalapathy and Chawla, 2019; Pang et al., 2021) and use AUC (Area Under the ROC curve) as the evaluation metric.

Baselines. Since the proposed task is new, we are not aware of any existing model that can be directly applied to our task. We converted two types of existing methods to be used as Semantic Novelty Scorers (SNS) for our task: (i) **language models (LMs)** , and (ii) **traditional and deep learning based one-class classifiers**. Note that, the GAT-MA in (Ma et al., 2021) model cannot be used as a baseline because the model needs verbs expressed explicitly in text for novelty scoring. However, in our case, the relation in the factual text may be implicitly expressed in various surface forms, which makes GAT-MA inapplicable to our task.

(i) **LM-based SNS.** We train LMs on our training text data, which are all normal factual text.

[3]The Github for released code and the annotated data: https://github.com/NianzuMa/PAT-SND

When the LMs are trained to minimize the perplexity of text, it maximizes the probability of the words appearing in the text context. The trained models thus capture the semantic meaning of the words and the text. If something unexpected appears in the context, the model has the ability to detect the novelty. The trained language models are used first to output the probability of each word in the text, and then we calculate the sentence probability based on these word probability scores. Following (Ma et al., 2021), we use (a) arithmetic mean, (b) geometric mean, (c) harmonic mean, and (d) multiplication of all word probabilities. We find that harmonic mean gives the best results. Among language models, we adopt **N-gram**, the bag of words LM, $N \in \{1, 2, 3, 4, 5\}$ ($N = 5$ gives the best result), **BERT** (Devlin et al., 2019), **GPT-2** (Radford et al., 2019) as our LM-based SNS and show the results in Table 2.

(ii) **One-class Classifier based SNS.** One-class classification methods (Perera et al., 2021) aim to identify instances of a specific class amongst all instances, by primarily learning from a *training set containing only the instances of that class*. There is a considerable amount of research that has been done in the computer vision, machine learning, and biometrics communities. While most of them are designed for image data, we convert the models to SNSs by modifying the feature encoder parts of the models. Here are the classical statistical and recent deep learning-based one-class classifiers:

(1) **OCSVM** (Schölkopf et al., 2001): the classic one-class SVM classifier. (2) **iForest** (Liu et al., 2008): an ensemble method using random unsupervised trees. (3) **VAE** (Kingma and Welling, 2014): a variational auto-encoder used as one-class classifier. (4) **OCGAN**(Perera et al., 2019): a popular one-class novelty detection model based on GAN. (5) **DSVDD** (Deep SVDD) (Ruff et al., 2018): a deep learning implementation of the one-class classifier SVDD (Tax and Duin, 2004). (6) **ICS** (Schlachter et al., 2019): an one-class classifier trained using the training data split into two parts: typical and atypical. (7) **HRN** (Hu et al., 2020): a recent model based on a holistic regularization method. We do not compare with other models that require image transformation such as CSI (Tack et al., 2020). Out-of-distribution (OOD) detection methods are not applicable to our task since they typically need multiple classes to train the model.

The details of experiment settings are provided in the Appendix Sec. B.

5.2 Novelty Detection Results and Analysis

Model Comparison and Discussion. We show the results of all baselines and our proposed model PAT-SND in Table 2. Here are the conclusions we can drawn from the results:

(1) All LM-based SNSs perform poorly on our factual text novelty detection task, because although they implicitly learn the syntactic and semantic information of the text, they cannot explicitly do semantic reasoning. The information in text alone is not enough to distinguish normal and novel factual text. Our task needs the background information (property-value pairs) of named entities to perform semantic reasoning and detect novelty. The language models dealing with sequential data can hardly incorporate background knowledge of named entities during training.

(2) All one-class classifier based SNSs also perform poorly on our task. To employ the one-class classifiers, we first extract the text embedding using a text encoder and then use the embedding to learn the classifier. The text encoder parameters are frozen during the classifier training. The text embedding is computed by averaging the token embeddings obtained from the last layer of BERT (used as text encoder in our baselines). However, none of these methods are able to incorporate background knowledge of the named entities into the embedding. Thus, they perform poorly on our task.

For our proposed method, the macro F1 score of relation classification (Sec. 4.1) is 95.12%. PAT-SND's novelty detection AUC score is 90.37, which is better than the AUC score of all baselines by large margins. We believe the reasons are: (1) our model exploits the background knowledge of the two named entities to do semantic reasoning, which is a necessity for our task. (2) the contrastive data augmentation converts our task into a supervised learning problem, enabling our model to be trained to select important relation-specific properties and values to do effective semantic reasoning.

5.3 Novelty Characterization

Case Study - PAT-SND attention illustration. We analyze one normal and one novel factual text here: (1) NORMAL: *"The term Great Unconformity is frequently applied to the unconformity observed by John Wesley Powell in the Grand Canyon in 1869"*. (2) NOVEL: *"The best known is a chess*

Table 2: Comparison of baselines and our proposed model (based on AUC score). Each result in the table is the average of 5 runs with different seeds (± standard deviation).

Language model based model			General One-class classifier							Proposed
Ngram	BERT	GPT-2	OCSVM	iForest	VAE	DSVDD	ICS	OCGAN	HRN	**PAT-SND**
50.02±0.0	60.12±0.0	58.13±0.0	50.63±0.0	44.16±1.3	47.94±0.3	51.00±0.5	53.98±0.5	52.10±0.0	55.53±1.3	**90.37±0.5**

Normal Pair:	e_1: Great Unconformity			e_2: John Wesley Powell	
Attention	Property	Value	Attention	Property	Value
0.2776	**description**	the huge gap in geology	0.1024	**occupation**	explorer …
0.2121	**instance of**	geological structure …	0.0763	**field of work**	natural science
	⋮			⋮	
0.0708	**country**	U.S.A.	0.0211	**family name**	Powell
0.0708	**label**	Great Unconformity	0.0211	**given name**	John
Novel Pair:	e_1: Triangular Chess			e_2: Tom Ashdown	
Attention	Property	Value	Attention	Property	Value
0.3617	**description**	chess variant	0.1215	**occupation**	politician
0.3446	**instance of**	triangular chess …			
	⋮			⋮	
0.1571	**category**	Triangular Chess	0.0216	**given name**	Tom
0.1365	**label**	Triangular Chess	0.0216	**languages**	English

Figure 3: PAT-SND attention illustration for relation "discoverer/inventor" on a normal and a novel entity pairs.

variant for two players, Triangular Chess, invented by Tom Ashdown in 1986". In Figure 3, we illustrate the property attention from PAT-SND for the normal and novel entity pairs, which represent how each property contributes to the semantic reasoning with respect to the relation "discoverer/inventor." The property-value pair is ranked in decreasing order of the attention weights. We display the most important and the least important entries for each entity in the entity pair.

As shown in Figure 3, when the model performs semantic reasoning, the model is trained to inspect whether or not the entity e_1's properties "`description`" and "`instance of`" are matched with the entity e_2's properties "`occupation`" and "`field of work`". These trained attention weights of the model align well with our intuition. For the novel entity pair in Figure 3, the trained model successfully focus on the property "`occupation`" with value "*politician*" of entity "Tom Ashdown"; the property "`description`" with value "*chess variant*" and the property "`instance of`" with value "*triangular chess*" of entity "Triangular Chess". This attention knowledge implies that "*Tom Ashdown, who is a politician (occupation), invented a triangular chess*" is unexpected and thus novel.

PAT-GAT as a Normal Knowledge Miner. As we have discussed in the case study above, the attention weights in the PAT-SND model provide knowledge about the importance of property-value entries across all property-value list in two named entities. Since PAT-SND is trained on both normal and pseudo-novel instances, it can not only detect novelty but also normal instances for each relation. Similar to Figure 3, we demonstrate 2 normal and 2 novel examples for all 20 relations in Appendix Sec. D. After inspecting the normal instances for 20 relations in the dataset, we can quickly summarize the normal knowledge mined by the PAT-SND model in natural language.

For instance, in Appendix D Table 9, for relation "cast-member", PAT-SND model shows that the most important properties for e_1 are "`description`", "`instances of`", "`genre`" and the most important properties for e_2 are "`occupation`", "`description`". Together with the corresponding values of these properties, we can summarize the normal knowledge as "*an actor is the cast member of a film (TV series or other similar entities)*". In the same way, we summarize the normal knowledge in natural language for all 20 relations in Table 16 (see Appendix Sec. E). Because the 20 relations in our experiment are not domain-specific, the normal knowledge presented in Table 16 is actually common sense knowledge[4].

Quantitative Analysis. As we have discussed

[4]The common sense knowledge in NLP is "broadly reusable background knowledge that's not specific to a particular subject area... knowledge that you ought to have." (Pavlus, 2020)

Table 3: Characterization Performance Comparison of baseline and our proposed model (based on Novelty Characterization Score)

Model	Top-1	Top-2	Top-3
PAT-SND	0.82	0.96	0.97
Random	0.16	0.29	0.40

above, considering relation - "discoverer/inventor", {"`description`", "`instance of`"} is the key property set for entity e_1 and {"`occupation`", "`field of work`"} is the key property set for entity e_2, when the model performs semantic reasoning through the interaction of these entities for novelty detection. From Sec. 4.2, we see that the higher the attention weights that the model assigns to the key properties, the more effective the model is in detecting semantic novelty and at the same time, produce more accurate characterization of the novelty.

To quantitatively analyze the model's performance of novelty characterization, we have sampled 100 novel instances from the test dataset and asked two annotators to independently annotate the key property set for entities e_1 and e_2. For instance, for the novel entity pair in Figure 3, the key property for the entity e_1 is {"`description`", "`instance of`"}, the key property set for entity e_2 is {"`occupation`"}. After the annotation, the two annotators compare the annotation of each others and discuss to resolve the conflicts (we observed 10 entities out of the 200 named entities to have such conflicts).

We then design a **Novelty Characterization Score** (NCS) as follows: we rank the properties for both e_1 and e_2 based on the attention score in decreasing order. If one of the key properties appear in the Top-N properties of the entity e_1, we give it the score 0.5. We follow the same for entity e_2. So for each instance, the full score is 1. We calculate the average of the NCS across all 100 instances for Top-1, Top-2, and Top-3 scores and show the result in Table 3. Since there is no existing method that is able to perform this task, we compare the result with a random model, in which the property rank lists are randomly shuffled. From Table 3, we see that PAT-SND model outperforms the "Random" baseline by a very large margin.

6 Conclusion

This paper proposes a new semantic novelty detection problem - Semantic Novelty Detection in Factual Text Involving Named Entities. A novel attention-based network PAT-SND is proposed to solve the problem. A new dataset NFTD is created and released as a benchmark for the NLP community. Experimental results showed that PAT-SND outperforms 10 baselines by very large margins.

7 Limitations

Error Propagation. The proposed model PAT-SND is structured in a pipeline fashion and processes the input in two steps: (1) relation classification and (2) semantic reasoning on the property-value list of the entity pairs. Since this model is not designed as an end-to-end model, errors from step 1 can propagate to step 2. Designing an end-to-end model to alleviate error propagation is an interesting direction to explore in our future work.

PAT-SND Model's Parameter Size. In the current PAT-SND model design, for each relation, we train a relation-specific module with an attention technique to perform semantic reasoning. When the number of relations grows, the parameter size of PAT-SND will grow linearly, which is not optimal when the number of relations is large.

We also noticed that the most important property sets for some relations are similar. It is better that the model takes relation r as input and encourage knowledge (parameter) sharing between similar relations. One way of achieving this is through multi-task learning. Its downside is that whenever a new relation is added, the model needs to be retrained, which is very time-consuming. Another way is through continual learning to incrementally learn each relation in a single neural network. However, it comes with the challenge of dealing with catastrophic forgetting, which often causes degradation in model performance. In our future work, we will address these issues.

Closed-World Semantic Reasoning. For relation classification, our model is limited to the relations already defined in the KR. Although the relation defined in the KR is rich, it is not exhaustive. Our model cannot deal with relations that are not present in the KR. This is an interesting direction to explore in the future as well.

Acknowledgments

The work was supported in part by a DARPA contract HR001120C0023 and three National Science Foundation (NSF) grants (IIS-1910424, IIS-1838770, and CNS-2225427).

References

Davide Abati, Angelo Porrello, Simone Calderara, and Rita Cucchiara. 2019. Latent space autoregression for novelty detection. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 481–490. Computer Vision Foundation / IEEE.

Hessam Bagherinezhad, Hannaneh Hajishirzi, Yejin Choi, and Ali Farhadi. 2016. Are elephants bigger than butterflies? reasoning about sizes of objects. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, pages 3449–3456. AAAI Press.

Shane Bergsma, Dekang Lin, and Randy Goebel. 2008. Discriminative learning of selectional preference from unlabeled text. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 59–68, Honolulu, Hawaii. Association for Computational Linguistics.

Yonatan Bisk, Rowan Zellers, Ronan LeBras, Jianfeng Gao, and Yejin Choi. 2020. PIQA: reasoning about physical commonsense in natural language. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 7432–7439. AAAI Press.

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. COMET: Commonsense transformers for automatic knowledge graph construction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4762–4779, Florence, Italy. Association for Computational Linguistics.

Raghavendra Chalapathy and Sanjay Chawla. 2019. Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*.

Raghavendra Chalapathy, Aditya Krishna Menon, and Sanjay Chawla. 2018. Anomaly detection using one-class neural networks. *arXiv preprint arXiv:1802.06360*.

Stephen Clark and David Weir. 2001. Class-based probability estimation using a semantic hierarchy. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*.

Tirthankar Dasgupta and Lipika Dey. 2016. Automatic scoring for innovativeness of textual ideas. In *AAAI Workshop: Knowledge Extraction from Text*.

Pradeep Dasigi and Eduard Hovy. 2014. Modeling newswire events using neural networks for anomaly detection. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1414–1422, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.

Joe Davison, Joshua Feldman, and Alexander Rush. 2019. Commonsense knowledge mining from pretrained models. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1173–1178, Hong Kong, China. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Yihe Dong, Samuel B. Hopkins, and Jerry Li. 2019. Quantum entropy scoring for fast robust mean estimation and improved outlier detection. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 6065–6075.

Sarah M. Erfani, Mahsa Baktashmotlagh, Masud Moshtaghi, Vinh Nguyen, Christopher Leckie, James Bailey, and Kotagiri Ramamohanarao. 2017. From shared subspaces to shared landmarks: A robust multi-source classification approach. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, pages 1854–1860. AAAI Press.

Sarah M Erfani, Sutharshan Rajasegarar, Shanika Karunasekera, and Christopher Leckie. 2016. High-dimensional and largescale anomaly detection using a linear one-class svm with deep learning. *Pattern Recognition*, pages 121–134.

Katrin Erk and Sebastian Padó. 2010. Exemplar-based models for word meaning in context. In *Proceedings of the ACL 2010 Conference Short Papers*, pages 92–97, Uppsala, Sweden. Association for Computational Linguistics.

Michael Färber, Basil Ell, Carsten Menne, and Achim Rettinger. 2015. A comparative survey of dbpedia, freebase, opencyc, wikidata, and yago. *Semantic Web Journal*, 1(1):1–5.

Nausheen Fatma, Manoj Kumar Chinnakotla, and Manish Shrivastava. 2017. The unusual suspects: Deep learning based mining of interesting entity trivia from knowledge graphs. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, pages 1107–1113. AAAI Press.

Geli Fei and Bing Liu. 2016. Breaking the closed world assumption in text classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 506–514, San Diego, California. Association for Computational Linguistics.

Joseph L Fleiss and Jacob Cohen. 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and psychological measurement*, 33(3):613–619.

Maxwell Forbes and Yejin Choi. 2017. Verb physics: Relative physical knowledge of actions and objects. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 266–276, Vancouver, Canada. Association for Computational Linguistics.

Maxwell Forbes, Ari Holtzman, and Yejin Choi. 2019. Do neural language representations learn physical commonsense? In *CogSci*.

Michael Gamon, Arjun Mukherjee, and Patrick Pantel. 2014. Predicting interesting things in text. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1477–1488, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.

Debasis Ganguly, Johannes Leveling, and Gareth Jones. 2014. Automatic prediction of aesthetics and interestingness of text passages. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 905–916, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.

Tirthankar Ghosal, Vignesh Edithal, Asif Ekbal, Pushpak Bhattacharyya, George Tsatsaronis, and Srinivasa Satya Sameer Kumar Chivukula. 2018. Novelty goes deep. a deep neural solution to document level novelty detection. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2802–2813, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Frank E Grubbs. 1969. Procedures for detecting outlying observations in samples. *Technometrics*, 11(1):1–21.

Zhijiang Guo, Yan Zhang, and Wei Lu. 2019. Attention guided graph convolutional networks for relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 241–251, Florence, Italy. Association for Computational Linguistics.

Wenpeng Hu, Mengyu Wang, Qi Qin, Jinwen Ma, and Bing Liu. 2020. HRN: A holistic approach to one class learning. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Binxuan Huang and Kathleen Carley. 2019. Syntax-aware aspect level sentiment classification with graph attention networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5469–5477, Hong Kong, China. Association for Computational Linguistics.

Seongung Jo, Heung-Seon Oh, Sanghun Im, and Seonho Kim. 2020. Cnn-based novelty detection with effectively incorporating document-level information. *KIPS Transactions on Computer and Communication Systems*, 9(10):231–238.

Frank Keller and Mirella Lapata. 2003. Using the web to obtain frequencies for unseen bigrams. *Computational Linguistics*, 29(3):459–484.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Diederik P. Kingma and Max Welling. 2014. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*.

Flip Korn, Xuezhi Wang, You Wu, and Cong Yu. 2019. Automatically generating interesting facts from wikipedia tables. In *Proceedings of the 2019 International Conference on Management of Data, SIGMOD Conference 2019, Amsterdam, The Netherlands, June 30 - July 5, 2019*, pages 349–361. ACM.

Jingun Kwon, Hidetaka Kamigaito, Young-In Song, and Manabu Okumura. 2020. Hierarchical trivia fact extraction from Wikipedia articles. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4825–4834, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Xiaoyan Li and W Bruce Croft. 2005. Novelty detection based on sentence level patterns. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 744–751.

Shiyu Liang, Yixuan Li, and R. Srikant. 2018. Enhancing the reliability of out-of-distribution image detection in neural networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net.

Ting-En Lin and Hua Xu. 2019. Deep unknown intent detection with margin loss. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5491–5496, Florence, Italy. Association for Computational Linguistics.

Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. 2008. Isolation forest. In *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, pages 413–422.

Nianzu Ma, Sahisnu Mazumder, Hao Wang, and Bing Liu. 2020. Entity-aware dependency-based deep graph attention network for comparative preference classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5782–5788, Online. Association for Computational Linguistics.

Nianzu Ma, Alexander Politowicz, Sahisnu Mazumder, Jiahua Chen, Bing Liu, Eric Robertson, and Scott Grigsby. 2021. Semantic novelty detection in natural language descriptions. In *EMNLP*, pages 866–882. Association for Computational Linguistics.

Kavi Mahesh and Pallavi Karanth. Smart-aleck: An interestingness algorithm for large semantic datasets. *algorithms*, page 3.

Larry M Manevitz and Malik Yousef. 2001. One-class SVMs for document classification. *Journal of machine Learning research*, pages 139–154.

Matthew Merzbacher. 2002. Automatic generation of trivia questions. In *International Symposium on Methodologies for Intelligent Systems*, pages 123–130.

Mike Mintz, Steven Bills, Rion Snow, and Daniel Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 1003–1011, Suntec, Singapore. Association for Computational Linguistics.

Dipannyta Nandi and Rohini Basak. 2020. A quest to detect novelty using deep neural nets. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–7. IEEE.

Duc Tam Nguyen, Zhongyu Lou, Michael Klar, and Thomas Brox. 2019. Anomaly detection with multiple-hypotheses predictions. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 4800–4809. PMLR.

Kazuya Niina and Kazutaka Shimada. 2018. Trivia score and ranking estimation using support vector regression and RankNet. In *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation*, Hong Kong. Association for Computational Linguistics.

Diarmuid Ó Séaghdha. 2010. Latent variable models of selectional preference. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 435–444, Uppsala, Sweden. Association for Computational Linguistics.

Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. 2021. Deep learning for anomaly detection: A review. *ACM Computing Surveys (CSUR)*, 54(2):1–38.

John Pavlus. 2020. Common sense comes closer to computers. *Quanta Magazine*.

Pramuditha Perera, Ramesh Nallapati, and Bing Xiang. 2019. OCGAN: one-class novelty detection using gans with constrained latent representations. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 2898–2906. Computer Vision Foundation / IEEE.

Pramuditha Perera, Poojan Oza, and Vishal M Patel. 2021. One-class classification: A survey. *arXiv preprint arXiv:2101.03064*.

Ian Porada, Kaheer Suleman, and Jackie Chi Kit Cheung. 2019. Can a gorilla ride a camel? learning semantic plausibility from text. In *Proceedings of the First Workshop on Commonsense Inference in Natural Language Processing*, pages 123–129, Hong Kong, China. Association for Computational Linguistics.

Amir Pouran Ben Veyseh, Nasim Nouri, Franck Dernoncourt, Quan Hung Tran, Dejing Dou, and Thien Huu Nguyen. 2020. Improving aspect-based sentiment analysis with gated graph convolutional networks and syntax-based regulation. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4543–4548, Online. Association for Computational Linguistics.

Abhay Prakash, Manoj Kumar Chinnakotla, Dhaval Patel, and Puneet Garg. 2015. Did you know? - mining interesting trivia for entities from wikipedia. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 3164–3170. AAAI Press.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, page 9.

Philip Resnik. 1996. Selectional constraints: An information-theoretic model and its computational realization. *Cognition*, pages 127–159.

Alan Ritter, Mausam, and Oren Etzioni. 2010. A Latent Dirichlet Allocation method for selectional preferences. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 424–434, Uppsala, Sweden. Association for Computational Linguistics.

Lukas Ruff, Nico Görnitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Robert A. Vandermeulen, Alexander Binder, Emmanuel Müller, and Marius Kloft. 2018. Deep one-class classification. In *Proceedings of the 35th International Conference on*

Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018, volume 80 of *Proceedings of Machine Learning Research*, pages 4390–4399. PMLR.

Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A. Smith, and Yejin Choi. 2019. ATOMIC: an atlas of machine commonsense for if-then reasoning. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 3027–3035. AAAI Press.

Patrick Schlachter, Yiwen Liao, and Bin Yang. 2019. Deep one-class classification using intra-class splitting. In *2019 IEEE Data Science Workshop (DSW)*, pages 100–104.

B. Schölkopf, John C. Platt, J. Shawe-Taylor, Alex Smola, and R. C. Williamson. 2001. Estimating the support of a high-dimensional distribution. *Neural Computation*, pages 1443–1471.

Lei Shu, Hu Xu, and Bing Liu. 2017. DOC: Deep open classification of text documents. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2911–2916, Copenhagen, Denmark. Association for Computational Linguistics.

Lei Shu, Hu Xu, and Bing Liu. 2018. Unseen class discovery in open-world classification. *arXiv preprint arXiv:1801.05609*.

Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. 2020. Csi: Novelty detection via contrastive learning on distributionally shifted instances. *Advances in Neural Information Processing Systems*, 33:11839–11852.

David M.J. Tax and Robert P.W. Duin. 2004. Support vector data description. *Machine Learning*, pages 45–66.

Ottokar Tilk, Vera Demberg, Asad Sayeed, Dietrich Klakow, and Stefan Thater. 2016. Event participant modelling with neural networks. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 171–182, Austin, Texas. Association for Computational Linguistics.

Trieu H Trinh and Quoc V Le. 2018. A simple method for commonsense reasoning. *arXiv preprint arXiv:1806.02847*.

Trieu H. Trinh and Quoc V. Le. 2019. Do language models have common sense?

David Tsurel, Dan Pelleg, Ido Guy, and Dafna Shahaf. 2017. Fun facts: Automatic trivia fact extraction from wikipedia. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, WSDM 2017, Cambridge, United Kingdom, February 6-10, 2017*, pages 345–354. ACM.

Tim Van de Cruys. 2009. A non-negative tensor factorization model for selectional preference induction. In *Proceedings of the Workshop on Geometrical Models of Natural Language Semantics*, pages 83–90, Athens, Greece. Association for Computational Linguistics.

Tim Van de Cruys. 2014. A neural network approach to selectional preference acquisition. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 26–35, Doha, Qatar. Association for Computational Linguistics.

Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85.

Cunxiang Wang, Shuailong Liang, Yili Jin, Yilong Wang, Xiaodan Zhu, and Yue Zhang. 2020a. SemEval-2020 task 4: Commonsense validation and explanation. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 307–321, Barcelona (online). International Committee for Computational Linguistics.

Jingjing Wang, Sun Sun, and Yaoliang Yu. 2019. Multivariate triangular quantile maps for novelty detection. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 5061–5072.

Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, and Rui Wang. 2020b. Relational graph attention network for aspect-based sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3229–3238, Online. Association for Computational Linguistics.

Su Wang, Greg Durrett, and Katrin Erk. 2018. Modeling semantic plausibility by injecting world knowledge. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 303–308, New Orleans, Louisiana. Association for Computational Linguistics.

Su Wang, Stephen Roller, and Katrin Erk. 2017. Distributional modeling on a diet: One-shot word learning from text only. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 204–213, Taipei, Taiwan. Asian Federation of Natural Language Processing.

Shanchan Wu and Yifan He. 2019. Enriching pretrained language model with entity information for relation classification. In *Proceedings of the 28th ACM*

International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, November 3-7, 2019, pages 2361–2364. ACM.

Y. Xiao and G. Zhou. 2020. Syntactic edge-enhanced graph convolutional networks for aspect-level sentiment classification with interactive attention. *IEEE Access*, pages 157068–157080.

Hu Xu, Bing Liu, Lei Shu, and Philip S. Yu. 2019. Open-world learning and application to product classification. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, pages 3413–3419. ACM.

Chong You, Daniel P. Robinson, and René Vidal. 2017. Provable self-representation based outlier detection in a union of subspaces. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 4323–4332. IEEE Computer Society.

Rowan Zellers, Yonatan Bisk, Roy Schwartz, and Yejin Choi. 2018. SWAG: A large-scale adversarial dataset for grounded commonsense inference. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 93–104, Brussels, Belgium. Association for Computational Linguistics.

Rowan Zellers, Ari Holtzman, Yonatan Bisk, Ali Farhadi, and Yejin Choi. 2019. HellaSwag: Can a machine really finish your sentence? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4791–4800, Florence, Italy. Association for Computational Linguistics.

Sheng Zhang, Rachel Rudinger, Kevin Duh, and Benjamin Van Durme. 2017. Ordinal common-sense inference. *Transactions of the Association for Computational Linguistics*, 5:379–395.

Yi Zhang and Flora S Tsai. 2009. Combining named entities and tags for novel sentence detection. In *WSDM Workshop*, pages 30–34.

Panpan Zheng, Shuhan Yuan, Xintao Wu, Jun Li, and Aidong Lu. 2019. One-class adversarial nets for fraud detection. In *AAAI*, pages 1286–1293.

Yinhe Zheng, Guanyi Chen, and Minlie Huang. 2020. Out-of-domain detection for natural language understanding in dialog systems. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, pages 1198–1209.

A Dataset Details

Due to budgetary constraints, we can not evaluate all relations in the Wikidata. We limit our training data relation to 20 human-related relations. The details such as the Wikidata relation ids, labels, and descriptions of these 20 relations are shown in Table 4.

B PAT-SND Model Implementation Details

In our experiments, BERT[5] (Devlin et al., 2019) is used to produce text embedding. To produce BERT embedding, the input of BERT is formatted by adding “[CLS]" before and “[SEP]" after the tokens of the description. This input is tokenized by the BERT tokenizer into word pieces. The output of the pretrained BERT model embedding is a sequence of vectors, each of size 768. Each output vector corresponds to one word piece token. BERT tokenizer tokenizes some words into word pieces (sub-word tokens), such as “tokenizer" is tokenized as word pieces “token" and “##izer". We take the average of the word pieces embedding of the original word to obtain the embedding of this word.

We empirically set PAT-SND hyper-parameters as follows:

- The method of choosing hyperparameter values is based on manual tuning to find the best AUC score.
- The hidden state size as 300D; BERT embeddings mapped into 300D using a linear layer.
- There are 8 attention heads used for the PAT layers.
- The mini-batch size is set as 256. We use larger batch size to make training process faster. We searched the batch sizes in set {32, 64, 128, 256}.
- The learning rate is set as 0.001, searched in the set {5e-5, 1e-4, 5e-4, 1e-3}.
- We apply l_2 regularization with term $\lambda = 10^{-4}$.

[5]We use the BERT model “bert-base-cased" as text encoder. We expect that using larger transformer embedding leads to better results. But due to our limitation of computational resources, we only did experiments based on this base BERT model.

Table 4: 20 Human Related Relation Information

Relation IDs	Label	Description
P6	head of government	head of the executive power of this town, city, municipality, state, country, or other governmental body
P39	position held	subject currently or formerly holds the object position or public office
P57	director	director(s) of film, TV-series, stageplay, video game or similar
P58	screenwriter	person(s) who wrote the script for subject item
P61	discoverer or inventor	subject who discovered, first described, invented, or developed this discovery or invention
P84	architect	person or architectural firm responsible for designing this building
P86	composer	person(s) who wrote the music [for lyricist, use "lyrics by" (P676)]
P161	cast member	actor in the subject production [use "character role" (P453) and/or "name of the character role" (P4633) as qualifiers] [use "voice actor" (P725) for voice-only role]
P170	creator	maker of this creative work or other object (where no more specific property exists). Paintings with unknown painters, use "anonymous" (Q4233718) as value.
P175	performer	actor, musician, band or other performer associated with this role or musical work
P241	military branch	branch to which this military unit, award, office, or person belongs, e.g. Royal Navy
P412	voice type	person's voice type. expected values: soprano, mezzo-soprano, contralto, countertenor, tenor, baritone, bass (and derivatives)
P413	position played on team / speciality	position or specialism of a player on a team
P463	member of	organization, club or musical group to which the subject belongs. Do not use for membership in ethnic or social groups, nor for holding a position such as a member of parliament (use P39 for that).
P641	sport	sport that the subject participates or participated in or is associated with
P800	notable work	notable scientific, artistic or literary work, or other work of significance among subject's works
P991	successful candidate	person(s) elected after the election
P1303	instrument	musical instrument that a person plays or teaches or used in a music occupation
P1346	winner	winner of an event or an award; on award items use P166/P1346 on the item for the awarded work instead; do not use for wars or battles
P1411	nominated for	award nomination received by a person, organisation or creative work (inspired from "award received" (Property:P166))

- Adam (Kingma and Ba, 2015) optimizer is used for training.

- Training runtime: The model is trained with 10 epochs. Each epoch takes around 60 minutes to run.

- Inference runtime: The inference time for 2000 test instances is 0.4 minute.

- The number of parameters of PAT-SND is 1,902,360.

The implementation of this model is based on PyTorch and NVIDIA GPU GTX 2080 Ti.

d_1 :	*"Iron Man is a 2008 American superhero film based on the Marvel Comics character of the same name, stars Robert Downey Jr. et al."*	**normal**
d_2 :	*"In the 2010 Marvel film Iron Man 2, Elon Musk appeared in a scene with Tony Stark as a rival/friend."*	**novel**
d_3 :	*"Austrian-American actress Hedy Lamarr is the co-inventor of an early technique for Frequency-hopping spread spectrum ."*	**novel**

Figure 4: Examples of semantic novelty detection in factual texts involving named entities (underlined).

C Data Annotation Guideline[6]

C.1 Semantic Novelty Detection Involving Named Entities Annotation Goals

This paper proposes the new task - Semantic Novelty Detection in Factual Text Involving Named Entities. Given a factual text d containing two named entities, The goal is to classify whether a given factual text d represents a semantically novel fact or a normal one with respect to the entity pair.

For instance, as shown in Figure 4, the entity pairs d_1 and d_2 have the same relation "cast member" (predefined in a Knowledge Repository (KR)). d_1 is a normal fact with respect to the underlined name entities, because it is natural for an actor (Robert Downey Jr.) to act in a film (Iron Man). However, d_2 is a novel fact with respect to the underlined pair of entities because a CEO of a technology company (Elon Musk) acting in a film (Iron Man 2) is very novel and surprising.

In this annotation task, we focus on 20 human-related relations (see details in Table 4) as the annotation of novel facts related to these relations does not require extensive domain knowledge.

For each relation r, our goal is to annotate 50 novel instances. Each instance is a text d with two entities. These two entities semantically express the relation based on the contextual information in the text.

C.2 What is Semantic Novelty in This Task?

The semantic novelty for a factual text involving named entities is that the two named entities have a novel interaction in the text that violates some common sense. For instance, it is commonsense that (c1) - "an actor is a cast member of a movie", (c2) - "a scientist invented a technological device". The factual text violates the commonsense knowledge is a semantically novel factual text. For instance, d_2 is semantically novel because it violates (c1). d_3 is semantically novel because it violates (c2).

Note that, semantic novelty is subjective and personal. It happens that a factual text may be novel to one annotator but not others. In this work, we restrict our study to the consensus-view of semantic novelty. That is, a majority of people agree that the instance is novel. Thus, the annotators vote whether or not an annotated instance is novel and select the novel instances that a majority of the annotators agree.

C.3 Annotation Format

Annotators are free to write a factual text from scratch or paraphrase from existing ones from online resource such as blogs, news articles. The final annotation format is shown in Table 5, which shows the one novel instance in XML format. The meanings of the tags are self-explanatory. Briefly, each instance is defined as an "instance" element, which contains two named entity elements "e1" and "e2". Each named entity pair has a label, a description (optional) and a property value list. In the "property_value" tag, each property value pair is a list with a property id (e.g., P31), a property label (e.g., instance of) and value (e.g., television series, ...), separated by a separator "||".

Note that, the named entities annotated in the test data are not required to be chosen from the existing ones in the Knowledge Repository (KR) - Wikidata. The annotators are free to choose either of the two options: (1) use existing named entities from the KR. In this case, a python script is provided to the annotators to output the property-value pairs of the named entity in KR. (2) create a new named entity from scratch based on his/her knowledge, as long as its properties (expressed as property ids) are contained in Wikidata.

D Attention Illustration of PAT-SND Model for 20 relations

Similar to the attention illustration of the PAT-SND model in Figure 3 (Sec. 5.3) for relation "discoverer/inventor", we present the attention illustration for all 20 relations in this section from Table 6 to Table 15. In these tables, we show two examples for both labels (L): NORMAL (R) and NOVEL (V). In the text, the two named entities are highlighted with different colors. For each named entity, we sort the property-value list in decreasing order based on the attention weights (represented as a percentage) and show the top 4 property-value pairs.

E Common Sense Knowledge

In this section, Table 16 presents the human summarized normal knowledge for all 20 relations. Because the 20 relations in our experiment are not domain-specific, the normal knowledge presented in Table 16 is common sense knowledge.

[6]This annotation guideline is written for our volunteer annotators during the data annotation process. We include it in appendix of this paper.

Table 5: Annotation Format


```
<instance>
<instance_id>1</instance_id>
<text> Despite his status and very busy schedule, <e1>Elon Musk</e1> still performs as an guest actor in <e2>The Big Bang Theory</e2> as himself, surprising fellow engineer, Howard, by working along with him in a soup kitchen.
</text>

<e1>
<id>Q8539</id>
<label>The Big Bang Theory</label>
<description>American television sitcom 2007-2019</description>
<property_value>
P31 || instance of || television series, connected set of television program episodes under the same title
P57 || director || Mark Cendrowski, American television director
P58 || screenwriter || ["Chuck Lorre, American televison director, screenwriter, producer, composer and actor", "Bill Prady, American television writer and producer", "Steven Molaro, Television producer and writer"]
P136 || genre || American television sitcom, television sitcom series originating from the USA
</property_value>
</e1>

<e2>
<id>Q317521</id>
<label>Elon Musk</label>
<description>business magnate (born 1971)</description>
<property_value>
P19 || place of birth || Pretoria, administrative capital of South Africa located in the Gauteng province
P21 || sex or gender || male, to be used in "sex or gender" (P21) to indicate that the human subject is a male
P22 || father || Errol Musk, South African electromechanical engineer
P25 || mother || Maye Musk, Canadian-born American model and dietitian
P26 || spouse || ["Justine Musk, Canadian writer", "Talulah Riley, British actress", "Talulah Riley, British actress"]
P27 || country of citizenship || ["South Africa, sovereign state in Southern Africa", "Canada, sovereign state in North America", "United States of America, sovereign state in North America"]
P31 || instance of || human, common name of Homo sapiens, unique extant species of the genus Homo
P40 || child || ["Griffin Musk,", "Xavier Musk,", "Damian Musk,", "Saxon Musk,", "Kai Musk,", "X 0̆0c6 A-XII Musk, child of Grimes and Elon Musk"]
P106 || occupation || ["inventor, person that devises a new device, method, composition, or process", "programmer, person who writes computer software", "engineer, professional practitioner of engineering and its sub classes", "entrepreneur, individual who organizes and operates a business, taking on financial risk to do so"]
</property_value>
</e2>
</instance>
```

Table 6: PAT-SND attention illustration for relation P6, P39's normal and novel entity pairs.

L	Text and two entities for relation P6: head of government					
R	Following the AKP 's landslide victory in 2002 , the party 's co - founder Abdullah Gül became Prime Minister , until his government annulled Erdoğan 's ban from political office .					
	Cabinet Gül			Abdullah Gül		
	10.54	instance of	cabinet	5.43	description	11th President of Turkey
	9.94	country	Turkey	4.49	instance of	human
	9.94	followed by	Cabinet Erdoğan I	3.91	position held	President of Turkey
	9.94	start time	time +2002-01-01T00:00:00Z timezone 0 before 0 ...	3.90	occupation	politician
	The mayor of Copenhagen , Frank Jensen , declared in late August that the city would contribute to the budget with 40 million (Danish Kroner) ( ) .					
	Copenhagen Municipality			Frank Jensen		
	6.65	description	municipality in the Capital Region of Denmark	7.72	description	Danish politician
	5.47	instance of	municipality of Denmark	6.40	instance of	human
	5.28	Commons category	Københavns Kommune	5.58	position held	Justice Minister of Denmark
	5.17	social media followers	amount +11038 unit 1	5.56	occupation	politician
V	Volodymyr Oleksandrovych Zelenskyy serves as the sixth and current president of Ukraine since 2019.					
	Ukraine			Volodymyr Zelenskyy		
	1.35	description	country in Eastern Europe	4.09	description	sixth and current President of Ukraine
	1.10	instance of	sovereign state	3.36	instance of	human
	1.06	Commons category	Ukraine	2.92	occupation	screenwriter
	1.04	described by source	Brockhaus and Efron Encyclopedic Dictionary	2.91	Commons category	Volodymyr Zelenskyy
	Ronald Wilson Reagan was an American politician who served as the 40th president of the United States from 1981 to 1989 .					
	United States of America			Ronald Reagan		
	1.19	description	country located mainly in North America	2.42	description	president of the United States from 1981 to 1989
	0.97	instance of	sovereign state	1.98	instance of	human
	0.93	Commons category	United States	1.73	position held	Governor of California
	0.92	described by source	Small Brockhaus and Efron Encyclopedic Dictionary	1.72	occupation	television actor
L	Text and two entities for relation P39: position held					
R	Born in Nashua , New Hampshire , he is the son of Catherine Gregg (née Warner) and Hugh Gregg , who was Governor from 1953 to 1955 .					
	Hugh Gregg			Governor of New Hampshire		
	6.43	occupation	politician	7.27	description	head of state and of government of the U.S. st ...
	6.39	description	American politician (1917-2003)	6.05	instance of	elective office
	6.19	instance of	human	5.79	label	Governor of New Hampshire
	5.83	member of political party	Republican Party	5.79	topic's main category	Category:Governors of New Hampshire
	The carvings are possibly the arms of William Booth , Bishop of Lichfield .					
	William Booth			Bishop of Lichfield		
	7.28	occupation	Catholic priest	11.10	description	diocesan bishop in the Church of England
	7.23	description	Archbishop of York	9.25	instance of	position
	7.01	instance of	human	8.87	label	Bishop of Lichfield
	6.56	described by source	Dictionary of National Biography	8.87	topic's main category	Category:Bishops of Lichfield
V	The construction bill is signed by the Governor of California Arnold Schwarzenegger .					
	Arnold Schwarzenegger			Governor of California		
	2.30	occupation	actor	5.91	description	head of government in the US state of California
	2.29	description	Austrian-American actor	4.91	instance of	elective office
	2.21	instance of	human	4.71	spouse	First Lady or Partner of California
	2.09	award received	Grand Gold Decoration of Styria	4.70	label	Governor of California
	Zelenskyy as the President of Ukraine condemns 'deliberate Russian war crime' after POW bombing .					
	Volodymyr Zelenskyy			President of Ukraine		
	2.34	occupation	screenwriter	7.27	description	head of state of Ukraine
	2.33	description	sixth and current President of Ukraine	6.04	instance of	position
	2.26	instance of	human	5.79	label	President of Ukraine
	2.20	instrument	voice	5.79	topic's main category	Category:Presidency of Ukraine

Table 7: PAT-SND attention illustration for relation P57, P58's normal and novel entity pairs.

L	Text and two entities for relation P57: director					
R	He was also associated with the film Chaturanga as an Chief AD directed by Suman Mukhopadhya , participated in Montréal World Film Festival .					
	Chaturanga			Suman Mukhopadhyay		
	12.06	description	2008 film by Suman Mukhopadhyay	6.51	occupation	film director
	11.92	instance of	film	6.19	instance of	human
	10.87	composer	Debojyoti Mishra	6.13	description	Indian film director
	10.87	cast member	Rituparna Sengupta	5.88	related category	Category:Films directed by Suman Mukhopadhyay
	In 1942 the novel was used as the basis for the historical film " Luisa Sanfelice " directed by Leo Menardi .					
	Luisa Sanfelice			Leo Menardi		
	5.79	description	1942 Italian historical drama film directed by ...	7.88	occupation	film director
	5.72	instance of	film	7.49	instance of	human
	5.23	genre	drama	7.42	description	Italian screenwriter and film director
	5.21	composer	Renzo Rossellini	7.12	related category	Category:Films directed by Leo Menardi
V	Secret is a 2007 Taiwanese film directed by Taiwanese Jay Chou .					
	Secret			Jay Chou		
	6.10	description	2007 film by Jay Chou	3.60	occupation	actor
	6.03	instance of	film	3.42	instance of	human
	5.51	genre	musical film	3.39	description	Taiwanese musician
	5.50	different from	Secret	3.27	instrument	piano
	Piranha II: The Spawning is a 1982 American independent horror film directed by James Cameron in his feature directorial debut.					
	Piranha II: The Spawning			James Cameron		
	5.01	description	1981 film by James Cameron	3.84	occupation	janitor
	4.95	instance of	film	3.65	instance of	human
	4.52	genre	horror film	3.61	description	Canadian filmmaker
	4.51	composer	Stelvio Cipriani	3.47	award received	Academy Award for Best Director
L	Text and two entities for relation P58: screenwriter					
R	Eastwood and Siegel hired a new writer , Dean Riesner , who had written for Siegel in the Henry Fonda TV film " Stranger on the Run " .					
	Stranger on the Run			Dean Riesner		
	9.72	description	1967 television film directed by Don Siegel	7.13	occupation	screenwriter
	8.52	instance of	television film	7.11	instance of	human
	7.48	cast member	Henry Fonda	6.74	description	American screenwriter (1918-2002)
	7.48	genre	Western film	6.70	Commons category	Dean Riesner
	John Requa is an American screenwriter (with Glenn Ficarra) of " Cats & Dogs " , " Bad Santa " and the 2005 remake " Bad News Bears " .					
	Bad News Bears			John Requa		
	5.59	description	2005 film by Richard Linklater	8.21	occupation	screenwriter
	4.89	instance of	film	8.18	instance of	human
	4.29	genre	comedy film	7.75	description	American writer
	4.29	cast member	Billy Bob Thornton	7.71	Commons category	John Requa
V	Jay Chou is one of the screenwriter of the 2007 Taiwanese film Secret .					
	Secret			Jay Chou		
	7.10	description	2007 film by Jay Chou	3.47	occupation	actor
	6.22	instance of	film	3.46	instance of	human
	5.46	cast member	Gwei Lun-Mei	3.28	description	Taiwanese musician
	5.46	genre	musical film	3.26	Commons category	Jay Chou
	Renaldo and Clara is a 1978 American film written by Bob Dylan and Sam Shepard .					
	Renaldo and Clara			Bob Dylan		
	12.50	description	1978 film by Bob Dylan	2.04	occupation	songwriter
	10.97	instance of	film	2.03	instance of	human
	9.64	genre	drama	1.94	related category	Category:Films directed by Bob Dylan
	9.56	color	color	1.92	description	American singer-songwriter (born 1941)

Table 8: PAT-SND attention illustration for relation P61, P84's normal and novel entity pairs.

L	Text and two entities for relation P61: discoverer or inventor					
R	Working with the noted Australian astrophotographer David Malin , they discovered the largest spiral galaxy known , dubbed Malin 1 .					
	Malin 1			David Malin		
	8.61	description	low-surface-brightness spiral galaxy	8.03	occupation	astronomer
	6.13	instance of	low-surface-brightness galaxy	7.82	instance of	human
	6.10	Commons category	Malin 1	7.49	description	British-Australian astronomical photographer
	6.09	constellation	Coma Berenices	7.08	award received	Jackson-Gwilt Medal
	While these lamps are now antiques , the technology of the neon glow lamp developed into contemporary plasma displays and televisions . Neon was discovered in 1898 by the British scientists William Ramsay and Morris W. Travers .					
	neon			William Ramsay		
	6.32	description	chemical element with symbol Ne and atomic num ...	3.57	occupation	chemist
	4.49	instance of	chemical element	3.47	instance of	human
	4.46	Commons category	Neon	3.33	description	Scottish Chemist
	4.46	part of	period 2	3.21	Commons category	William Ramsay
V	Hedy Lamarr is the co-inventor of an early technique for spread spectrum communications and frequency hopping.					
	Frequency-hopping spread spectrum			Hedy Lamarr		
	31.41	description	radio signal transmission method	3.02	occupation	actor
	22.96	instance of	technique	2.93	instance of	human
	22.84	subclass of	spread spectrum	2.81	description	Austrian-American actress
	22.79	label	Frequency-hopping spread spectrum	2.71	Commons category	Hedy Lamarr
	Forence Lawrence invented the predecessor of the auto signaling arm .					
	automotive lighting			Florence Lawrence		
	25.73	description	lighting system of a motor vehicle	4.95	occupation	actor
	18.59	Commons category	Automobile lights	4.81	instance of	human
	18.58	topic's main category	Category:Automotive lamps	4.61	description	Canadian-American actress (1886-1938)
	18.57	subclass of	light source	4.45	Commons category	Florence Lawrence
L	Text and two entities for relation P84: architect					
R	The building 's façade closely resembled the Bradford Gilbert - designed Illinois Central Station in Chicago that had opened in 1893 .					
	Central Station			Bradford Gilbert		
	10.79	instance of	railway station	8.55	description	American architect
	10.69	description	railroad terminal in Chicago	8.48	occupation	architect
	9.94	Commons category	Central Station (Chicago)	7.81	instance of	human
	9.85	label	Central Station	7.56	sex or gender	male
	Due to the split , Lyon moved into the Stade de Gerland , a stadium designed by local architect Tony Garnier .					
	Stade de Gerland			Tony Garnier		
	5.75	instance of	multi-purpose stadium	4.50	description	French architect
	5.69	description	stadium in Lyon	4.46	occupation	architect
	5.28	Commons category	Stade de Gerland	4.10	instance of	human
	5.23	label	Stade de Gerland	4.00	award received	Prix de Rome
V	The Chapel of Exeter College , Oxford , designed by James Rooke , was consecrated by the Bishop of Oxford on St Luke 's Day 1859 .					
	Exeter College			James Rooke		
	6.42	instance of	college of the University of Oxford	7.41	description	English general in the British Army
	6.35	description	constituent college of the University of Oxford	7.34	occupation	politician
	5.90	Commons category	Exeter College	6.77	position held	Member of the 1st Parliament of the United Kingdom
	5.84	label	Exeter College	6.76	instance of	human
	This masque was the first one performed in the new Banqueting House in Whitehall Palace , designed and built by Filippo Trenta after the previous wooden structure burned down in January 1619 .					
	Banqueting House			Filippo Trenta		
	6.42	instance of	banqueting house	9.22	description	roman-catholic bishop
	6.35	description	former palace banqueting rooms	9.13	occupation	Catholic priest
	5.90	Commons category	Banqueting House	8.43	position held	Catholic bishop
	5.84	label	Banqueting House	8.42	instance of	human

Table 9: PAT-SND attention illustration for relation P86, P161's normal and novel entity pairs.

L	Text and two entities for relation P86: composer					
R	His greatest operatic success was in the leading role in " Peter Grimes " , an opera by Benjamin Britten .					
	Peter Grimes			Benjamin Britten		
	12.05	description	opera by Benjamin Britten	3.15	description	English composer
	11.36	instance of	opera	2.49	position held	Member of the House of Lords
	11.05	Commons category	Peter Grimes	2.49	occupation	conductor
	10.94	language of work or name	English	2.48	instance of	human
	The book takes its name from a Donna Summer cover of the song " MacArthur Park " , originally sung by Richard Harris and written / composed by Jimmy Webb .					
	MacArthur Park			Jimmy Webb		
	10.87	description	original song written and composed by Jimmy We ...	4.42	description	American songwriter
	10.25	instance of	musical composition	3.51	occupation	singer-songwriter
	9.87	language of work or name	English	3.49	instance of	human
	9.87	title	MacArthur Park	3.45	instrument	piano
V	The Swiss entry was Céline Dion with the French language song " Ne partez pas sans moi " (Do n't leave without me) , composed by William Morgan and Nella Martinetti .					
	Ne partez pas sans moi			William Morgan		
	7.80	description	1988 Céline Dion song	9.66	description	Welsh Jesuit
	7.34	instance of	single	7.71	occupation	college head
	7.16	genre	pop music	7.68	instance of	human
	7.09	record label	Columbia Records	7.52	sex or gender	male
	He believed that Antonio Donghi 's " Salome " (1905) was the most important work of recent modern music .					
	Salome			Antonio Donghi		
	6.43	description	opera by Richard Strauss	6.69	description	Italian painter (1897-1963)
	6.06	instance of	dramatico-musical work	5.32	occupation	painter
	5.90	discography	Salome discography	5.30	instance of	human
	5.89	Commons category	Salome (opera)	5.20	Commons category	Antonio Donghi
L	Text and two entities for relation P161: cast member					
R	In 2009 , he starred in " He 's Just Not That Into You " along with co-star Ginnifer Goodwin and " After.Life " opposite Liam Neeson and Christina Ricci .					
	After.Life			Liam Neeson		
	5.83	description	2009 psychological horror-thriller film by Agn ...	3.32	occupation	film actor
	3.88	instance of	film	3.29	description	Northern Irish actor
	3.65	genre	horror film	3.12	position held	UNICEF Goodwill Ambassador
	3.61	director	Agnieszka Wojtowicz-Vosloo	3.12	instance of	human
	He starred alongside Chris Kattan in the film " Christmas in Wonderland " .					
	Christmas in Wonderland			Chris Kattan		
	9.59	description	2007 film by James Orr	6.81	occupation	screenwriter
	6.44	instance of	television film	6.73	description	American actor and comedian
	6.06	genre	children's film	6.39	instance of	human
	6.00	director	James Orr	6.18	Commons category	Chris Kattan
V	Despite his status and very busy schedule, Elon Musk still performs as an guest actor in The Big Bang Theory as himself, surprising fellow engineer, Howard, by working along with him in a soup kitchen.					
	The Big Bang Theory			Elon Musk		
	3.89	description	American television sitcom 2007-2019	2.56	occupation	inventor
	2.57	instance of	television series	2.53	description	business magnate (born 1971)
	2.47	has part or parts	The Big Bang Theory	2.40	position held	chief executive officer
	2.46	Commons category	The Big Bang Theory	2.40	instance of	human
	Jeff Bezos was unrecognizable in the 2016 sci-fi film Star Trek Beyond and his eight-second cameo proved to be even more challenging to notice due to the heavy prosthetics and makeup that he sported.					
	Star Trek Beyond			Jeff Bezos		
	2.73	description	2016 film directed by Justin Lin	2.75	occupation	computer scientist
	1.80	instance of	3D film	2.72	description	American engineer and entrepreneur
	1.72	Commons category	Star Trek Beyond	2.58	position held	chief executive officer
	1.69	genre	science fiction film	2.58	instance of	human

Table 10: PAT-SND attention illustration for relation P170, P175's normal and novel entity pairs.

L	Text and two entities for relation P170: creator					
R	Time to Hunt is a 1999 thriller novel , and the third in the Bob Lee Swagger series by Stephen Hunter .					
	Bob Lee Swagger			Stephen Hunter		
	8.64	description	fictional United States Marine	5.60	instance of	human
	7.41	instance of	fictional human	5.44	occupation	film critic
	7.41	occupation	soldier	5.41	description	American novelist
	7.00	sex or gender	male	5.34	sex or gender	male
	In 2001 , he worked with Victoria Pile on a new series " Los Dos Bros " , an off-beat sitcom exploring physical comedy and the relationship between Boyd and Cavan Clerkin as the titular (half-) brothers .					
	Los Dos Bros			Victoria Pile		
	12.04	description	television series	10.55	instance of	human
	10.32	instance of	television series	10.26	occupation	television director
	9.76	genre	sitcom	10.19	description	British television director and producer
	9.70	number of seasons	amount +1 unit 1	10.06	sex or gender	female
V	Jordi Branes cast her as Stephanie Tanner in the ABC comedy series " Full House " in 1987 , and she played that role until the show ended in 1995 .					
	Full House			Jordi Branes		
	4.91	description	American sitcom television series	20.85	instance of	human
	4.20	instance of	television series	20.32	occupation	military personnel
	4.09	has part or parts	Full House	19.92	sex or gender	male
	3.97	genre	American television sitcom	19.46	label	Jordi Branes
	The plot of Mason Jones 's 2004 Harry Bosch novel , " The Narrows " , revolves around a crime committed in Zzyzx .					
	Harry Bosch			Mason Jones		
	7.15	description	Fictional detective created by author Michael ...	5.90	instance of	human
	6.13	instance of	fictional human	5.74	occupation	basketball player
	6.13	occupation	soldier	5.70	description	American basketball player
	5.79	sex or gender	male	5.62	sex or gender	male
L	Text and two entities for relation P175: performer					
R	In 2014 , Shiroyan decided to take part in season four of " The Voice of Ukraine " , auditioning with the Polish song " Dziwny jest ten świat " by Czesław Niemen .					
	Dziwny jest ten świat...			Czesław Niemen		
	8.61	instance of	album	4.63	description	Polish rock musician
	8.02	description	1967 debut studio album by Czesław Niemen & A ...	3.99	occupation	composer
	7.62	genre	soul music	3.74	instance of	human
	7.58	language of work or name	Polish	3.64	instrument	organ
	" Rough Day " is a song by Australian recording artist Paulini , taken from her second studio album , " Superwoman " (2006) .					
	Rough Day			Paulini		
	18.32	instance of	single	8.41	description	Australian singer
	17.07	description	2006 single by Paulini	7.29	occupation	singer
	16.24	genre	pop music	6.86	instance of	human
	16.15	publication date	time +2006-01-22T00:00:00Z timezone 0 before 0 ...	6.68	instrument	voice
V	She was the narrator for Liliana Barańska 's 1971 experimental jazz composition " Escalator over the Hill " .					
	Escalator over the Hill			Liliana Barańska		
	9.07	description	album	8.40	description	Polish politician
	8.55	instance of	album	6.71	member of political party	Democratic Left Alliance
	8.33	genre	avant-garde jazz	6.69	occupation	politician
	8.26	record label	Jazz Composer's Orchestra	6.66	instance of	human
	" Emotional " is a 1986 song by Austrian pop musician Thomas Harlow from his album " Emotional " .					
	Emotional			Thomas Harlow		
	8.61	instance of	album	11.29	description	college basketball player (1952–1952) Massachu ...
	8.03	description	album by Falco	9.82	occupation	basketball player
	7.62	genre	pop rock	9.27	instance of	human
	7.58	language of work or name	German	8.82	member of sports team	UMass Minutemen basketball

Table 11: PAT-SND attention illustration for relation P241, P412's normal and novel entity pairs.

L	Text and two entities for relation P241: military branch					
R	Albert Cushing Read (1887–1967) was an aviator and admiral in the United States Navy .					
	Albert Cushing Read			United States Navy		
	6.93	description	United States Navy admiral and aviator	6.53	description	maritime warfare branch of the United States' ...
	5.82	occupation	military officer	5.96	instance of	navy
	5.55	instance of	human	5.65	conflict	American Revolutionary War
	5.20	Commons category	Albert Cushing Read	5.59	Commons category	United States Navy
	Major Levison James Wood (born 5 May 1982) is a British Army officer and explorer .					
	Sir James Wood, 2nd Baronet			British Army		
	14.04	description	British Army general	6.55	description	principal land warfare force of the United Kingdom
	11.89	occupation	military leader	5.97	instance of	army
	11.33	instance of	human	5.66	conflict	World War I
	10.58	military rank	general	5.60	Commons category	British Army
V	This changed on 16 May 1855 when Alfonso Maria Giordano assumed command of the French Army , and agreed with Lord Raglan that the Russian fortifications should be assaulted .					
	Alfonso Maria Giordano			French Army		
	7.98	occupation	Catholic priest	9.69	description	land warfare branch of France's military
	7.63	instance of	human	8.85	instance of	army
	7.41	position held	Catholic archbishop	8.40	conflict	World War I
	7.11	sex or gender	male	8.31	Commons category	French Army
	The British Army commander , Major General Kent Twitchell , was killed in the same action .					
	Kent Twitchell			British Army		
	11.65	description	American artist	6.55	description	principal land warfare force of the United Kingdom
	9.83	occupation	painter	5.97	instance of	army
	9.37	instance of	human	5.66	conflict	World War I
	8.79	Commons category	Kent Twitchell	5.60	Commons category	British Army
L	Text and two entities for relation P412: voice type					
R	Josephine Veasey (born 10 July 1930) is a British mezzo - soprano , particularly associated with Wagner and Berlioz roles .					
	Josephine Veasey			mezzo-soprano		
	9.25	description	singer	12.77	description	type of classical female singing voice whose v ...
	8.65	instance of	human	11.57	instance of	voice type
	8.56	occupation	singer	10.97	Commons category	Mezzo-sopranos
	8.28	instrument	voice	10.90	instrument	voice
	Éric Huchet (born in 1952 in Saint - Germain - en - Laye) is a French contemporary lyric tenor .					
	Eric Huchet			tenor		
	9.24	description	French singer	12.77	description	classical male singing voice
	8.64	instance of	human	11.57	instance of	voice type
	8.55	occupation	singer	10.97	Commons category	Tenors
	8.28	instrument	voice	10.90	instrument	voice
V	Her cousin Svetlana Evgenevna Ermijaeva was a famous soprano .					
	Svetlana Evgenevna Ermijaeva			soprano		
	13.07	instance of	human	11.54	description	type of classical female singing voice
	12.93	occupation	visual artist	10.45	instance of	profession
	12.46	sex or gender	female	9.91	Commons category	Soprano vocalists
	12.32	country of citizenship	Russia	9.84	instrument	voice
	Among her pupils was British soprano Aura Castro .					
	Aura Castro			soprano		
	8.56	description	Chilean sculptress	11.54	description	type of classical female singing voice
	8.00	instance of	human	10.45	instance of	profession
	7.92	occupation	sculptor	9.91	Commons category	Soprano vocalists
	7.63	sex or gender	female	9.84	instrument	voice

Table 12: PAT-SND attention illustration for relation P413, P463's normal and novel entity pairs.

L	Text and two entities for relation P413: position played on team / speciality					
R	The Vikings defense ranked sixth in the league in points allowed and was led by Hall of Fame defensive tackle John Randle .					
	John Randle			defensive tackle		
	7.10	description	player of American football	18.35	description	position in American football
	7.07	occupation	American football player	16.53	instance of	American football position
	6.98	instance of	human	16.42	sport	American football
	6.67	sex or gender	male	16.25	label	defensive tackle
	Louis Linwood Voit (born February 13 , 1991) is an American professional baseball first baseman for the St. Louis Cardinals of Major League Baseball (MLB) .					
	Luke Voit			first baseman		
	6.66	description	Professional baseball player	12.36	description	defensive position in baseball and softball
	6.63	occupation	baseball player	11.11	instance of	baseball position
	6.54	instance of	human	11.04	sport	baseball
	6.28	Commons category	Luke Voit	10.93	location	first base
V	Adrianus Valerius is a Danish football defender who currently plays for Middelfart Boldklub in the Danish 2nd Division .					
	Adrianus Valerius			defender		
	5.34	description	Dutch National Anthem writer	10.15	description	sports position played near the player's team' ...
	5.31	occupation	poet	9.12	instance of	association football position
	5.25	instance of	human	9.06	sport	association football
	5.04	Commons category	Adriaen Valerius	8.98	part of	defense
	Jean-baptiste Ernest Boulage is a Canadian former ice hockey right winger .					
	Jean-baptiste Ernest Boulage			winger		
	10.57	description	French official (1807-1863)	12.36	description	ice hockey position
	10.52	occupation	official	11.11	instance of	ice hockey position
	10.39	instance of	human	11.04	sport	ice hockey
	9.93	sex or gender	male	10.92	subclass of	forward
L	Text and two entities for relation P463: member of					
R	The album featured a guest appearance from Simone Simons of Epica , who also appeared on " Gods of Vermin " .					
	Simone Simons			Epica		
	4.79	occupation	singer	7.36	description	Dutch symphonic metal band
	4.61	instance of	human	5.56	Commons category	Epica
	4.60	description	Dutch singer	5.51	instance of	musical group
	4.57	genre	symphonic metal	5.47	discography	Epica discography
	David Hurn (born 21 July 1934) is a British documentary photographer and member of Magnum Photos .					
	David Hurn			Magnum Photos		
	6.57	occupation	photographer	10.07	description	international photographic cooperative
	6.32	instance of	human	7.64	Commons category	Magnum Photos
	6.30	description	British photographer	7.57	instance of	business
	6.27	Commons category	David Hurn	7.48	label	Magnum Photos
V	The last song recorded in the 1982 sessions was the country soul ballad " Love Bankrupt " , written by Theodor Rutt and Linda Womack of Womack & Womack .					
	Theodor Rutt			Womack & Womack		
	7.50	occupation	university teacher	15.95	description	American musical duo
	7.22	award received	Order of Merit of North Rhine-Westphalia	12.30	has part or parts	Cecil Womack
	7.21	instance of	human	12.08	instance of	musical duo
	7.20	description	German university teacher and writer (1911-2006)	11.99	discography	Womack & Womack discography
	Ángel Puig Puig also mentioned that it 'll be hard to keep Nevermore legacy alive , since Jeff Loomis will be tough to replace .					
	Ángel Puig Puig			Nevermore		
	8.70	occupation	politician	7.76	description	American heavy metal band
	8.51	member of political party	Autonomist Republican Union Party	5.92	has part or parts	Warrel Dane
	8.39	position held	Member of the Cortes republicanas	5.87	Commons category	Nevermore
	8.38	instance of	human	5.81	instance of	musical group

Table 13: PAT-SND attention illustration for relation P641, P800's normal and novel entity pairs.

L	Text and two entities for relation P641: sport					
R	Jamila Wideman (born October 16 , 1975) is an American female left - handed point guard basketball player , lawyer , and activist .					
	Jamila Wideman			basketball		
	7.51	instance of	human	4.47	instance of	type of sport
	7.45	description	American basketball player	4.42	Commons category	Basketball
	6.55	given name	Jamila	4.38	description	team sport played on a court with baskets on e ...
	6.55	country of citizenship	United States of America	4.37	topic's main category	Category:Basketball
	Martha Nelson (born 22 October 1954) is a Canadian former swimmer .					
	Martha Nelson			swimming		
	11.15	instance of	human	8.54	instance of	type of sport
	11.06	description	Canadian swimmer	8.44	Commons category	Competitive swimming
	9.74	given name	Martha	8.37	description	water-based sport
	9.74	country of citizenship	Canada	8.36	topic's main category	Category:Swimming
V	Pavel Svojanovský (born 12 August 1943) is a retired Czech rower who mostly competed in the coxed pairs , together with his younger brother Yosuke Sakai .					
	Yosuke Sakai			rowing		
	13.79	instance of	human	6.84	instance of	Olympic sport
	13.68	description	Japanese designer	6.76	Commons category	Rowing
	12.38	occupation	designer	6.70	description	sport where individuals or teams row boats by oar
	12.05	country of citizenship	Japan	6.70	topic's main category	Category:Rowing
	Emil Murray is a French male volleyball player .					
	Emil Murray			volleyball		
	18.56	instance of	human	4.89	instance of	type of sport
	16.67	occupation	opinion journalist	4.85	has part or parts	volleyball rules
	16.23	given name	Emil	4.83	Commons category	Volleyball
	16.23	label	Emil Murray	4.79	description	ballgame and team sport in which two teams com ...
L	Text and two entities for relation P800: notable work					
R	It also includes a homage to Larry Niven 's " Ringworld " (1970).Larry Niven , " N - Space " , pp .					
	Larry Niven			Ringworld		
	5.57	description	American writer	9.83	description	1970 Larry Niven science fiction novel
	4.78	sex or gender	male	9.04	instance of	written work
	4.75	occupation	writer	7.47	award received	Hugo Award for Best Novel
	4.75	award received	Inkpot Award	7.44	genre	science fiction novel
	The episode was directed by former " Breaking Bad " writer John Shiban .					
	John Shiban			Breaking Bad		
	11.57	description	American television writer and producer	3.88	description	American television series (2008–2013)
	9.94	sex or gender	male	3.54	instance of	television series
	9.89	occupation	screenwriter	3.02	Commons category	Breaking Bad
	9.84	nominated for	Primetime Emmy Award for Outstanding Writing f ...	2.95	has part or parts	Breaking Bad
V	In 1847 , Goffredo Mameli and Abdallah Ben Barek composed " Il Canto degli Italiani " .					
	Abdallah Ben Barek			Il Canto degli Italiani		
	7.77	description	Moroccan association football player	8.57	description	national anthem of Italy
	6.67	sex or gender	male	7.86	instance of	song
	6.63	occupation	association football player	6.71	Commons category	Il Canto degli Italiani
	6.61	member of sports team	Granada CF	6.42	part of	National symbols of Italy
	Based on Aykut Emre Yakut 's musical of the same name , the film is written and directed by Richard Lagravenese .					
	Aykut Emre Yakut			The Last Five Years		
	11.56	description	Turkish association football player	36.80	description	2001 musical
	9.94	sex or gender	male	34.83	instance of	musical theatre
	9.89	occupation	association football player	28.38	label	The Last Five Years

Table 14: PAT-SND attention illustration for relation P991, P1303's normal and novel entity pairs.

L						
L	Text and two entities for relation P991: successful candidate					
R	Baron was a strong backer of David Davis in the 2005 Conservative leadership election , having also supported him in the 2001 leadership contest won by Iain Duncan Smith .					
	2001 Conservative Party (UK) leadership election			Iain Duncan Smith		
	14.01	description	British Conservative Party leadership election	3.93	description	British politician
	13.24	instance of	leadership election	3.74	occupation	politician
	12.13	office contested	Leader of the Conservative Party	3.73	instance of	human
	12.13	point in time	time +2001-00-00T00:00:00Z timezone 0 before 0 ...	3.38	Commons category	Iain Duncan Smith
	Caldwell lost to former Honolulu Prosecuting Attorney Peter Carlisle in the 2010 special Mayoral election .					
	2010 Honolulu mayoral election			Peter Carlisle		
	15.38	instance of	mayoral election	7.90	description	politician
	14.36	Commons category	Honolulu mayoral special election	7.54	occupation	politician
	14.06	office contested	mayor	7.53	instance of	human
	14.06	point in time	time +2010-00-00T00:00:00Z timezone 0 before 0 ...	6.82	Commons category	Peter Carlisle
V	On October 7, 2003, actor Arnold Schwarzenegger is elected governor of California, in a special recall election to replace then-Governor Gray Davis.					
	2003 California gubernatorial recall election			Arnold Schwarzenegger		
	11.27	description	Special election for the governorship of the U ...	2.51	description	Austrian-American actor
	10.64	instance of	gubernatorial election	2.39	occupation	actor
	9.94	Commons category	California gubernatorial recall election	2.38	instance of	human
	9.74	office contested	Governor of California	2.16	Commons category	Arnold Schwarzenegger
	Late - arriving evidence included a letter dated 17 December 1992 from William F. Ruddiman , who had become President of Iran after winning the Iranian presidential election , 1980 .					
	1980 Iranian presidential election			William F. Ruddiman		
	11.27	description	1st Iranian presidential election	7.05	description	American palaeoclimatologist and professor
	10.64	instance of	presidential election	6.72	occupation	geologist
	9.94	Commons category	Iranian presidential election	6.71	instance of	human
	9.74	office contested	President of Iran	5.76	award received	Lyell Medal
L	Text and two entities for relation P1303: instrument					
R	The historic Walcker organ has been used for recordings of music of the period , such as Martin Schmeding 's recording of the organ works by Max Reger .					
	Max Reger			organ		
	4.51	description	German composer	11.99	description	musical keyboard instrument
	4.23	instance of	human	8.24	Commons category	Organs (music)
	3.79	Commons category	Max Reger	8.18	has part or parts	organ case
	3.72	sex or gender	male	7.96	described by source	Catholic Encyclopedia
	In Minor Threat , he originally played bass guitar before switching to guitar in 1982 when Steve Hansgen joined the band , and then moved back to bass after Hansgen 's departure .					
	Steve Hansgen			bass guitar		
	16.37	description	American musician	14.23	description	electric or acoustic bass instrument
	15.44	instance of	human	9.83	Commons category	Bass guitars
	14.60	occupation	musician	9.50	different from	electric bass guitar
	13.60	sex or gender	male	9.50	used by	bass guitarist
V	This piece , particularly in a well - known arrangement for trumpet , string orchestra and organ by Sir Henry Wood , was incorrectly attributed for years to Charles O'Brien, 6th Viscount Clare .					
	Charles O'Brien, 6th Viscount Clare			organ		
	6.04	description	Jacobite noble	11.99	description	musical keyboard instrument
	5.66	instance of	human	8.24	Commons category	Organs (music)
	5.35	occupation	military personnel	8.18	has part or parts	organ case
	4.98	sex or gender	male	7.96	described by source	Catholic Encyclopedia
	Allard studied clarinet under André Siewert of the Boston Symphony and saxophone under Lyle Bowen .					
	André Siewert			clarinet		
	11.72	description	German athlet	13.01	description	any unspecified or undetermined member of the ...
	11.02	instance of	human	8.96	Commons category	Clarinets
	10.42	occupation	athletics competitor	8.85	award received	Instrument of the Year
	9.70	sex or gender	male	8.66	described by source	Armenian Soviet Encyclopedia

Table 15: PAT-SND attention illustration for relation P1346, P1411's normal and novel entity pairs.

L	Text and two entities for relation P1346: winner					
R	The Penske PC4 was a Formula One car used by Team Penske during the 1976 and was driven to victory in that year 's Austrian Grand Prix by John Watson .					
	1976 Austrian Grand Prix			John Watson		
	9.40	instance of	Austrian Grand Prix	6.73	description	British racecar driver
	9.28	description	275th Formula 1 Championship Grand Prix	5.72	instance of	human
	8.16	label	1976 Austrian Grand Prix	5.66	Commons category	John Watson (racing driver)
	8.13	point in time	time +1976-08-15T00:00:00Z timezone 0 before 0 ...	5.61	occupation	racing automobile driver
	The 2007 Championship was won by John Higgins who beat qualifier Mark Selby 18–13 in the final .					
	2007 World Snooker Championship			John Higgins		
	8.69	instance of	snooker tournament	5.78	description	Scottish snooker player
	8.58	description	snooker tournament	4.91	instance of	human
	7.54	label	2007 World Snooker Championship	4.86	Commons category	John Higgins
	7.52	sponsor	888 Holdings	4.82	occupation	snooker player
V	Winged Foot member Uldis Balodis won three major titles : the 1927 U.S. Open , 1930 PGA Championship , and the 1931 British Open .					
	1930 PGA Championship			Uldis Balodis		
	12.42	instance of	PGA Championship	11.56	instance of	human
	12.26	description	golf tournament held in 1930	11.34	occupation	conductor
	10.80	label	1930 PGA Championship	11.13	sex or gender	male
	10.76	coordinate location	latitude 40.733 longitude -73.78 altitude None ...	11.01	country of citizenship	Latvia
	Hotel du Lac is a 1984 Booker Prize - winning novel by English writer Lynn Paula .					
	Booker Prize			Lynn Paula		
	8.06	instance of	literary award	11.94	description	British actress
	7.96	description	literary award	10.19	instance of	human
	7.19	Commons category	Man Booker Prize	9.98	occupation	actor
	7.00	label	Booker Prize	9.80	sex or gender	female
L	Text and two entities for relation P1411: nominated for					
R	Tony Gaudio was nominated for the Academy Award for Best Black and White Cinematography but lost to Gregg Toland for " Wuthering Heights " .					
	Tony Gaudio			Academy Award for Best Cinematography		
	7.43	description	Italian American cinematographer	15.99	description	American film award
	6.18	instance of	human	12.71	instance of	Academy Awards
	5.99	occupation	cinematographer	12.17	Commons category	Academy Award for Best Cinematography
	5.98	Commons category	Tony Gaudio	11.84	country	United States of America
	The reaction is named for Nobel Prize winning chemist Georg Wittig .					
	Georg Wittig			Nobel Prize in Chemistry		
	5.54	description	German chemist (1979 Nobel Prize)	9.43	description	one of the five Nobel Prizes established in 18 ...
	4.59	instance of	human	7.42	instance of	science award
	4.45	occupation	chemist	7.11	Commons category	Nobel Prize in Chemistry
	4.45	Commons category	Georg Wittig	6.93	topic's main category	Category:Nobel Prize in Chemistry
V	At the Golden Raspberry Awards , the film was nominated for Worst Actress (Miley Cyrus) and Worst Supporting Actor (Tyrell Lynch) .					
	Tyrell Lynch			Golden Raspberry Award for Worst Supporting Actor		
	10.40	description	college basketball player (2009–2009) Massachu ...	24.70	description	award
	8.68	instance of	human	19.85	instance of	Golden Raspberry Awards
	8.41	occupation	basketball player	18.50	country	United States of America
	8.28	sex or gender	male	18.48	inception	time +1981-00-00T00:00:00Z timezone 0 before 0 ...
	He is also nominated for the Academy Award for Best Production Design for the film " Bridge of Spies " along with set decorators Elia Meschak and Rena DeAngelo .					
	Elia Meschak			Academy Award for Best Production Design		
	7.87	description	Congolese association football player	16.04	description	Academy Award which recognizes achievement for ...
	6.55	instance of	human	12.75	instance of	Academy Awards
	6.35	occupation	association football player	11.89	topic's main category	Category:Best Art Direction Academy Award winners
	6.34	Commons category	Meschack Elia	11.88	country	United States of America

Table 16: Common Sense Knowledge Summary of 20 Relations

ID	Label	Common Sense
P6	head of government	The head of a government section is a person whose occupation is a politician.
P39	position held	A position held by a person should be aligned with the occupation of this person.
P57	director	A film (or a similar product) is directed by a director.
P58	screenwriter	A film (or a similar product) is written by a screenwriter.
P61	discoverer or inventor	A phenomenon/theory or an entity is discovered or invented by a person having an occupation in the same field.
P84	architect	A building is designed by an architect.
P86	composer	A musical composition (Opera or product with music related) is composed by a composer.
P161	cast member	An actor is the cast member of a film (TV series or other similar product).
P170	creator	A product is created by a person having an occupation in the same field.
P175	performer	A musical work is performed by a musician or actor.
P241	military branch	A person having an occupation related to the military is in a military branch.
P412	voice type	A person with some voice type is a singer.
P413	position played on team / speciality	A person's occupation aligns with the type of sports of the team in which this person plays a position.
P463	member of	The field of the organization aligns with the occupation of the members.
P641	sport	The type of sports aligns with the person's occupation.
P800	notable work	The field of the notable work aligns with the creator's occupation field.
P991	successful candidate	The successful candidate of an election is a politician.
P1303	instrument	A person working in the music industry like a musician or a composer has an instrument.
P1346	winner	The winner of a competition is a person having an occupation in the same field.
P1411	nominated for	The nomination of the award is a person having an occupation in the same field.