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

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). One example of such work is the one by Ma et al. (2021), which requires factual reasoning over text as compared to normal one). This task is more fine-grained and involves novel scene descriptions (e.g., “A person walks a chicken that of (1), which has been studied extensively. (2) While 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.

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

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
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 entrepreneur) ....]. However, not all property-value pairs are 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 $D_{tr} = \{d_1, d_2, \ldots, d_n\}$, with each $d_i \in D_{tr}$, 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 $D_{tr}$, our goal is to build a model $F$ to score the semantic novelty of a test factual text $d'$ having a pair of entities $(e_1', e_2')$ with respect to $D_{tr}$, $\mathcal{K}$, and pair $(e_1^i, e_2^i)$, i.e., classifying $d'$ into one of the classes \{NORMAL, NOVEL\}. As $F$ 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 (Property ATention network for Semantic Novelty Detection) 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 Property ATention 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 (Novel Factual Text Detection) dataset. We leverage a distant supervision technique (Mintz et al., 2009) with the Wikipedia² 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.,

²https://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; Ö Séaghdha, 2010; Van de Cruys, 2009, 2014; Dasigi and Hovy, 2014; Tilk et al., 2016). These works differ from ours as we focus on fine-grained semantic novelty detection.

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]]” 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    | 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: \(\mathcal{K}R = (\mathcal{E}, \mathcal{R}, \mathcal{T})\), 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 \(e = \text{"Elon Mask"}\) 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 $K$).

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 $K$ 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 $K$ in step 2 [Sec. 4.2]. As our training data $D_{tr}$ consists of only NORMAL class examples (as discussed in Sec. 1), it’s not possible to train SNS solely with $D_{tr}$. 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 $D_{tr}$ 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 $D_{tr}$, where each $d_i \in D_{tr}$ is labelled with true relation label $r$ between the entity pair through the distant supervision technique.

4.2 Triple Semantic Novelty Scoring (SNS)

Let $B_1 = \{(p^1_i, v^1_i)|1 \leq i \leq l\}$ and $B_2 = \{(p^2_j, v^2_j)|1 \leq j \leq m\}$ be the background knowledge obtained for $e_1$ and $e_2$ respectively from KB $K$ (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.

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 $\{(p_i, v_i)|1 \leq i \leq N\}$ and a relation $r$ as input and outputs a weighted value vector $h^{pat}$ to be used for inference. $p_i$ and the corresponding $v_i$ are fed to PAT as feature vectors $p_1, v_1$ 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 $[CLS]$, instance of, $[SEP]$ 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 \( \{ p_i \}_{i=1}^N \) are fed one by one through a relation-specific linear layer, and a \( \text{relu} \) non-linearity function and a softmax function are used to obtain the attention weights \( \{ \alpha_{tr} \}_{i=1}^N \) over \( \{ p_i \}_{i=1}^N \) with respect to \( r \). Next, the weights are used to weigh the corresponding \( \{ v_i \}_{i=1}^N \) to obtain \( h^{out} \). The processing for a given \( r \) is summarized:

\[
\begin{align*}
g^r_k &= \text{relu}(p_i W^k_r \cdot b^k_r) \\
\alpha^r_k &= \frac{\exp(g^r_k)}{\sum_{k=1}^{K} \exp(g^r_k)} \\
h^{out} &= \sum_{i=1}^{N} \left( \frac{1}{K} \sum_{k=1}^{K} \alpha^r_k v^i_k \right)
\end{align*}
\]  

(1)

where \( K \) is the total number of attention heads and \( W^k_r, b^k_r \) are relation-specific weight and bias for the \( k \)-th attention head. \( \alpha^r_k \) is the \( k \)-th attention weight between \( r \) and \( p_i \). Overall, the processing of inputs in PAT is denoted as \( h^{out} = PAT(p_i, V, r; \Theta_r) \), where \( P = \{ p_1, p_2, \ldots, p_N \} \in \mathbb{R}^{N \times F} \) is the property matrix, \( V = \{ v_1, v_2, \ldots, v_N \} \in \mathbb{R}^{N \times F} \), is the value matrix and \( \Theta_r \) 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 \( P_1, V_1 \) from \( B_1 \) and \( P_2, V_2 \) from \( B_2 \) and feed them to PAT for relation \( \hat{r} \) as follows:

\[
\begin{align*}
h^{out}_1 &= PAT(P_1, V_1, \hat{r}; \Theta_r) \\
h^{out}_2 &= PAT(P_2, V_2, \hat{r}; \Theta_r) \\
h^{out}_k &= [h^{out}_1; h^{out}_2]
\end{align*}
\]

Next, a relation-specific feed-forward layer is used to project \( h^{out}_k \) into a semantic novelty score as \( S(\hat{r}) = (h^{out}_k W^f + b_f) \), where \( \hat{r} \) 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 \( S(\hat{r}) \) directly in our experiments (Sec. 5).

4.3 Training

Let \( T_{tr} \) be the set of all triples (labelled as NORMAL class) extracted from the examples in \( D_{tr} \).

To train SNS, we use KB \( K \) to help generate contrastive examples (triples) by corrupting the triples in \( T_{tr} \), 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 \( \tau_1 \in T_{tr} \), the generator \( G_{\text{contrastive}}(\tau_1) \) randomly samples an entity \( e' \) from KB \( K \) to replace either \( e_1 \) or \( e_2 \) in \( \tau_1 \). After corruption, \( \tau'_1 \) is formed from \( \tau_1 \), where \( \tau'_1 = (e', r, e_2) \) or \( \tau'_1 = (e_1, r, e') \). For example, given \( \tau_1 = \text{(The Big Bang Theory, cast-member, Johnny Galecki)} \) as a NORMAL triple in \( T_{tr} \), a pseudo-novel triple generated by \( G_{\text{contrastive}}(\tau_1) \) would be \( \tau'_1 = \text{(The Big Bang Theory, cast-member, Warren Buffett)} \). During the training of SNS, we dynamically generate one pseudo-novel triple for each NORMAL triple in \( T_{tr} \) in every training epoch.

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

\[
\mathcal{L} = \sum_{\tau \in T_{tr}} \sum_{\tau' \in T_{tr}'} \max\{S(\tau') - S(\tau) + 1, 0\}
\]

(3)

where, \( T_{tr}' \) is the set of pseudo-novel triples generated from \( T_{tr} \). \( \mathcal{L} \) encourages the score \( S(\tau) \) of the NORMAL triple \( \tau \) to be higher than \( S(\tau') \) of a pseudo-novel triple \( \tau' \).

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 draw from the results:

1. (i) 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. (ii) 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 | PAT-SND |
| 50.02±0.00 60.12±0.00 58.13±0.00 | 50.63±0.00 44.16±1.13 47.94±0.03 51.00±0.03 53.98±0.03 52.10±0.00 55.53±1.13 | 90.37±0.85 |

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

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.

**Quantitative Analysis.** As we have discussed...
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 multitask 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, Hananneh 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:1901.05407.

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 Moshtagh, 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.

Seongun 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, Hitotaka 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.

Shiyou 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 selection 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, Shaobai 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
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, AAI 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 Tsunel, 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 Yaoiang 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 pre-trained language model with entity information for relation classification. In Proceedings of the 28th ACM
A Dataset Details

Due to budgetary constraints, we cannot 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)] and/or "name of the character role" (P4633) as qualifiers]          |
| P161        | cast member               | actor in the subject production [use "character role" (P453) and/or "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 / specialty | position or specialism of a player on a team                                                                                                               |
| P463        | member of organization 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.
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 commonsense. 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.
Table 5: Annotation Format

```xml
<instance>
<instance_id>1</instance_id>
<text> Despite his status and very busy schedule, <e1>Elon Musk</e1> still performs as a 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 television 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 Óòc6 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>
```
| L | Text and two entities for relation P6: head of government |
|---|---|
| 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. |

| R | |
|---|---|
| Cabinet Gül | Abdullah Gül |
| 10.54 instance of | cabinet |
| 9.94 country | Turkey |
| 9.94 followed by | Cabinet Erdoğan I |
| 9.94 start time | time +2002-01-01T00:00:00Z timezone 0 before 0... |

| V | |
|---|---|
| The mayor of Copenhagen, | Frank Jensen, declared in late August that the city would contribute to the budget with 40 million (Danish Kroner) ( ). |

| L | Text and two entities for relation P39: position held |
|---|---|
| Born in Nashua, New Hampshire, he is the son of Katherine Gregg (née Warner) and Hugh Gregg, who was Governor of New Hampshire from 1953 to 1955. |

| R | |
|---|---|
| Hugh Gregg | Governor of New Hampshire |
| 6.43 occupation | politician |
| 6.39 description | American politician (1917-2003) |
| 6.19 instance of | Republican Party |

| V | |
|---|---|
| The carvings are possibly the arms of William Booth, Bishop of Lichfield. |

| L | Text and two entities for relation P39: position held |
|---|---|
| The construction bill is signed by the Governor of California Arnold Schwarzenegger. |

| R | |
|---|---|
| Arnold Schwarzenegger | Governor of California |
| 2.30 occupation | actor |
| 2.29 description | Austrian-American actor |
| 2.21 instance of | human |

| V | |
|---|---|
| The President of Ukraine Volodymyr Zelenskyy condemns ‘deliberate Russian war crime’ after POW bombing. |

| L | Text and two entities for relation P39: position held |
|---|---|
| Zelenskyy as the President of Ukraine condemns ‘deliberate Russian war crime’ after POW bombing. |

| R | |
|---|---|
| Volodymyr Zelenskyy | President of Ukraine |
| 2.34 occupation | screenwriter |
| 2.33 description | sixth and current President of Ukraine |
| 2.26 instance of | human |
| 2.20 instrument | voice |
Table 7: PAT-SND attention illustration for relation P57, P58’s normal and novel entity pairs.

| Text and two entities for relation P57: director | Text and two entities for relation P58: screenwriter |
|-----------------------------------------------|---------------------------------------------------|
| He was also associated with the film *Chaturanga* as a Chief AD directed by Suman Mukhopadhyay, participated in Montréal World Film Festival. | Eastwood and Siegel hired a new writer, Dean Riesner, who had written for Siegel in the Henry Fonda TV film *Stranger on the Run*. |
| 12.06 description 2008 film by Suman Mukhopadhyay | 9.72 description 1967 television film directed by Don Siegel |
| 11.92 instance of film | 8.52 instance of television film |
| 10.87 composer Debojyoti Mishra | 7.48 cast member Henry Fonda |
| 10.87 cast member Rituparna Sengupta | 7.48 genre Western film |
| In 1942 the novel was used as the basis for the historical film *Luisa Sanfelice* directed by Leo Menardi. | John Requa is an American screenwriter (with Glenn Ficarra) of *Cats & Dogs*, *Bad Santa* and the 2005 remake of *Bad News Bears*. |
| Secret is a 2007 Taiwanese film directed by Taiwanese Jay Chou. | "Bad News Bears". |
| Piranha II: The Spawning is a 1982 American independent horror film directed by James Cameron in his feature directorial debut. | |
| Piranha II: The Spawning | |
| 5.79 description 1942 Italian historical drama film directed by ... | 5.01 description 1981 film by James Cameron |
| 5.72 instance of film | 4.95 instance of film |
| 5.23 genre drama | 4.52 genre horror film |
| 5.21 composer Renzo Rossellini | 4.51 composer Stelvio Cipriani |
| Secret | 5.50 different from Secret |
| 6.10 description 2007 film by Jay Chou | 6.03 instance of film |
| 5.51 genre musical film | 5.50 instance of film |
| 5.50 different from Secret | 5.00 instance of musical film |
| 6.12 description 2005 film by Richard Linklater | 4.98 instance of film |
| 4.99 genre comedy film | 4.29 cast member Billy Bob Thornton |
| 4.29 cast member Billy Bob Thornton | 4.29 cast member John Requa |
| 7.10 description 2007 film by Jay Chou | 7.10 description 2005 film by Richard Linklater |
| 6.22 instance of film | 6.22 instance of film |
| 5.46 cast member Gwei Lun-Mei | 5.46 cast member Gwei Lun-Mei |
| 5.46 genre musical film | 5.46 genre musical film |
| 5.56 color | 5.56 color |

| Jay Chou is one of the screenwriter of the 2007 Taiwanese film *Secret*. | John Requa is an American screenwriter (with Glenn Ficarra) of *Cats & Dogs*, *Bad Santa* and the 2005 remake of *Bad News Bears*. |
|-----------------------------------------------|---------------------------------------------------|
| 7.10 description 2007 film by Jay Chou | 7.59 description 2005 film by Richard Linklater |
| 6.22 instance of film | 6.22 instance of film |
| 5.46 cast member Gwei Lun-Mei | 5.46 cast member Gwei Lun-Mei |
| 5.46 genre musical film | 5.46 genre musical film |

9243
Table 8: PAT-SND attention illustration for relation P61, P84’s normal and novel entity pairs.

| Text and two entities for relation P61: discoverer or inventor | Malin 1 | Malin 1 |
|-----------------|---------|---------|
| Working with the noted Australian astrophotographer David Malin, they discovered the largest spiral galaxy known, dubbed Malin 1. | | |
| 8.61 description low-surface-brightness spiral galaxy | 8.03 occupation astronomer | |
| 6.13 instance of low-surface-brightness galaxy Malin 1 | 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-Guill Medall | |
| 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. | | |
| 6.32 description chemical element with symbol Ne and atomic number 10 | 3.57 occupation chemist | |
| 4.49 instance of chemical element Ne | 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 | |
| Hedy Lamarr is the co-inventor of an early technique for frequency hopping spread spectrum communications and frequency hopping. | | |
| Frequency hopping spread spectrum | | |
| 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 | |
| Florence Lawrence invented the predecessor of the automotive lighting system of a motor vehicle. | | |
| Automotive lighting | | |
| 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 | |
| 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 | |
| The Chapel of Exeter College, Oxford, designed by James Rook, was consecrated by the Bishop of Oxford on St Luke’s Day 1859. | | |
| Exeter College | | James Rook |
| 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 | |

9244
| L | Text and two entities for relation P86: composer |
|---|---|
| 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 |

| R | Text and two entities for relation P161: normal and novel entity pairs. |
|---|---|
| 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 | Text and two entities for relation P161: cast member |
|---|---|
| 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 |
|---|---|
| Time to Hunt is a 1999 thriller novel, and the third in the Bob Lee Swagger series by Stephen Hunter. |
| **8.64** description | fictional United States Marine |
| **7.41** instance of | fictional human |
| **7.41** occupation | soldier |
| **7.00** sex or gender | male |

| R | Text and two entities for relation P170: creator |
|---|---|
| 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. |
| **8.61** description | television series |
| **10.32** instance of | television series |
| **9.76** genre | sitcom |
| **9.70** number of seasons | amount +1 unit 1 |

| L | Text and two entities for relation P175: performer |
|---|---|
| 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. |
| **9.61** description | American sitcom television series |
| **4.20** instance of | television series |
| **4.09** has part or parts | Full House |
| **3.97** genre | American television sitcom |
| **7.59** sex or gender | male |

| R | Text and two entities for relation P175: performer |
|---|---|
| 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. |
| **8.61** instance of | album |
| **8.02** description | 1967 debut studio album by Czesław Niemen & A... |
| **7.62** genre | soul music |
| **7.58** language of work or name | Polish |

| L | Text and two entities for relation P175: performer |
|---|---|
| She was the narrator for Liliana Barański's 1971 experimental jazz composition Escalator over the Hill. |
| **8.07** description | album |
| **8.55** instance of | album |
| **8.33** genre | avant-garde jazz |
| **8.26** record label | Jazz Composer's Orchestra |

| R | Text and two entities for relation P175: performer |
|---|---|
| "Emotional" is a 1986 song by Austrian pop musician Thomas Harlow from his album Emotional. |
| **8.61** instance of | album |
| **8.03** description | album by Falco |
| **7.62** genre | pop rock |
| **7.58** language of work or name | German |

| L | Text and two entities for relation P175: performer |
|---|---|
| Rough Day is a song by Australian recording artist Paulini, taken from her second studio album, 'Superwoman' (2006). |
| **18.32** instance of | single |
| **17.07** description | 2006 single by Paulini |
| **16.24** genre | pop music |
| **16.15** publication date | time+2006-01-22T00:00:00Z timezone 0 before 0... |

| R | Text and two entities for relation P175: performer |
|---|---|
| "Emotional" is a 1986 song by Austrian pop musician Thomas Harlow from his album Emotional. |
| **8.61** instance of | album |
| **8.03** description | album by Falco |
| **7.62** genre | pop rock |
| **7.58** language of work or name | German |
Table 11: PAT-SND attention illustration for relation P241, P412’s normal and novel entity pairs.

| Text and two entities for relation P241: military branch |
|---------------------------------------------------------|
| **Albert Cushing Read** (1887–1967) was an aviator and admiral in the United States Navy. |
| **Sir James Wood,** 2nd Baronet, was a British Army officer and explorer. |
| **Alfonso Maria Giordano** assumed command of the French Army, and agreed with Lord Raglan that the Russian fortifications should be assaulted. |

| Text and two entities for relation P412: voice type |
|--------------------------------------------------|
| **Josephine Veasey** (born 10 July 1930) is a British mezzo-soprano, particularly associated with Wagner and Berlioz roles. |
| **Éric Huchet** (born in 1952 in Saint-Germain-en-Laye) is a French contemporary lyric tenor. |
| **Her cousin** Svetlana Evgenevna Ermiyjeva was a famous soprano. |

| Occupation and instance of various entities |
|--------------------------------------------|
| **United States Navy admiral and aviator** |
| **United States’ maritime warfare branch of the military** |
| **navy** |
| **American Revolutionary War** |
| **Navy** |
| **Maritime warfare branch of the United States’ military** |
| **army** |
| **World War I** |
| **American artist** |
| **principal land warfare force of France’s military** |
| **French singer** |
| **classical male singing voice** |
| **voice** |
| **Site of World War I** |
| **Soprano** |
| **Soprano vocalists** |

| Description and instance of various categories |
|-----------------------------------------------|
| **United States Navy admiral** |
| **United States’ maritime warfare branch of the military** |
| **navy** |
| **American Revolutionary War** |
| **Navy** |
| **Maritime warfare branch of the United States’ military** |
| **army** |
| **World War I** |
| **American artist** |
| **principal land warfare force of France’s military** |
| **French singer** |
| **classical male singing voice** |
| **voice** |
| **Site of World War I** |
| **Soprano** |
| **Soprano vocalists** |

9247
| L | The Vikings defense ranked sixth in the league in points allowed and was led by Hall of Fame defensive tackle John Randle. |
|---|---|
| 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 | instance of | American football |
| 6.67 | sex or gender | male | 16.25 | label | defensive tackle |
| R | Louis Linwood Voit (born February 13, 1991) is an American professional baseball player for the St. Louis Cardinals of Major League Baseball (MLB). |
| 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. |
|---|---|
| 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 | Adrianus Valerius | 8.98 | part of | defense |
| 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 | The album featured a guest appearance from Simone Simons of Epica, who also appeared on “Gods of Vermin”. |
|---|---|
| 4.79 | occupation | singer | 7.36 | description | Dutch symphonic metal band |
| 4.61 | instance of | human | 5.56 | instance of | Epica |
| 4.60 | description | Dutch singer | 5.51 | instance of | musical group |
| 4.57 | genre | symphonic metal | 5.47 | instance of | discography |
| R | David Hurn (born 21 July 1934) is a British documentary photographer and member of 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 Ángel Puig Puig and Linda Womack of 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 | description | Womack & Womack discography |

| L | Ángel Puig Puig also mentioned that it’ll be hard to keep Nevermore legacy alive, since Jeff Loomis will be tough to replace. |
|---|---|
| 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 |
| Relation: P641, P800 | Normal Entity Pairs | Novel Entity Pairs |
|-----------------------|---------------------|------------------|
| **Jamila Wideman** | Born October 16, 1975 | American female left-handed point guard basketball player, lawyer, and activist. |
| Instance of: Jamila Wideman | Human | Basketball |
| Description: Jamila | American basketball player | Commons category: Basketball |
| 6.55 Country of citizenship: Jamila | United States of America | Team sport played on a court with baskets on either end. |
| **Martha Nelson** | Born October 22, 1954 | Canadian former basketball player, lawyer, and activist. |
| Instance of: Martha Nelson | Human | Basketball |
| Description: Martha | Canadian basketball player | Commons category: Basketball |
| 7.45 Description: Jamila | American basketball player | Description: American basketball player |
| 6.45 Description: Jamila | Canadian basketball player | Description: Canadian basketball player |
| **Pavel Svojanovský** | Born August 12, 1943 | Retired Czech rower who mostly competed in the coxed pairs, together with his younger brother Yosuke Sakai. |
| Instance of: Pavel Svojanovský | Human | Olympic sport |
| Description: Pavel | Czech rower | Commons category: Rowing |
| 6.9 Country of citizenship: Pavel | United States of America | Description: Team sport where individuals or teams row boats by oar. |
| **Emil Murray** | Born August 12, 1943 | French male volleyball player. |
| Instance of: Emil Murray | Human | Olympic sport |
| Description: Emil | French volleyball player | Commons category: Volleyball |
| 6.55 Description: Emil | Canadian volleyball player | Description: Team sport where individuals or teams hit a ball with a net. |

---

| Relation: P800 | Notable Works | Novel Works |
|----------------|---------------|-------------|
| **Larry Niven** | Ringsworld | 1970 Larry Niven science fiction novel |
| Description: Larry | American writer | Description: American television writer and producer. |
| 5.78 Sex or gender: Larry | Male | Description: American television series (2008–2013) |
| 7.55 Occupation: Larry | Writer | Description: American television series (2008–2013) |
| 4.75 Award received: Larry | Inkpot Award | Description: Hugo Award for Best Novel |
| **John Shiban** | The episode was directed by former Breaking Bad writer. |
| Instance of: John | Human | 1970 Larry Niven science fiction novel |
| Description: John | Writer | Description: American television writer and producer. |
| 5.78 Sex or gender: John | Male | Description: American television series |
| 7.55 Occupation: John | Writer | Description: American television series |
| **Abdallah Ben Barek** | In 1847, Gozzredo Mameli and Abdallah Ben Barek composed Il Canto degli Italiani. |
| Description: Abdallah | American association football player | Description: National anthem of Italy |
| 7.77 Sex or gender: Abdallah | Male | Description: National anthem of Italy |
| 7.6 Country of citizenship: Abdallah | Morocco | Description: National anthem of Italy |
| **Aykut Emre Yakut** | Based on Aykut Emre Yakut’s musical of the same name, the film is written and directed by Richard Lagravenese. |
| Description: Aykut | Turkish association football player | Description: 2001 musical |
| 11.56 Sex or gender: Aykut | Male | Description: 2001 musical |
| 9.45 Occupation: Aykut | Association football player | Description: 2001 musical |

---

9249
Table 14: PAT-SND attention illustration for relation P991, P1303’s normal and novel entity pairs.

| Text and two entities for relation P991: successful candidate | Text and two entities for relation P1303: instrument |
|--------------------------------------------------------------|---------------------------------------------------|
| **R** 14.01 description British Conservative Party leadership election | 14.01 description British Conservative Party leadership election |
| 13.24 instance of leadership election | 13.24 instance of leadership election |
| 12.13 instance of Human | 12.13 instance of Human |
| 12.13 point in time time +2001-00-00T00:00:00Z | 12.13 point in time time +2001-00-00T00:00:00Z |
| **L** Baron was a strong backer of David Davis in the 2005 Conservative leadership election, having also supported him in the 2001 Conservative Party (UK) leadership election | **V** Caldwell lost to former Honolulu Prosecuting Attorney Peter Carlisle in the 2010 special Mayoral election. |
| **R** Iain Duncan Smith | **V** Peter Carlisle |
| **L** 3.93 description British politician | 7.90 description politician |
| 3.74 occupation politician | 7.54 occupation politician |
| 3.73 instance of Human | 7.53 instance of Human |
| 3.38 Commons category Iain Duncan Smith | 6.82 Commons category Peter Carlisle |
| **L** 2001 Conservative Party (UK) leadership election | Caldwell was a strong backer of David Davis in the 2005 Conservative leadership election, having also supported him in the 2001 Conservative Party (UK) leadership election, having also supported him in the 2001 Conservative Party (UK) leadership election, having also supported him in the 2001 Conservative Party (UK) leadership election |
| **V** 2010 Honolulu mayoral election | 2010 Honolulu mayoral election |
| **L** 15.38 instance of mayor | 14.06 point in time time +2010-00-00T00:00:00Z |
| 14.36 Commons category Honolulu mayoral special election | 12.13 point in time time +2010-00-00T00:00:00Z |
| **V** 2003 California gubernatorial recall election | 2010 special Mayoral election |
| **L** 11.27 description Special election for the governorship of the U... | 14.06 instance of mayor |
| 10.64 instance of gubernatorial election | 9.74 instance of Governor of California |
| 9.94 Commons category California gubernatorial recall election | 2.16 Commons category Arnold Schwarzenegger |
| **V** 1980 Iranian presidential election | 1980 Iranian presidential election |
| **L** 11.27 description 1st Iranian presidential election | 11.27 description Special election for the governorship of the U... |
| 10.64 instance of presidential election | 10.64 instance of gubernatorial election |
| 9.94 Commons category Iranian presidential election | 9.94 Commons category Iranian presidential election |
| 9.74 office contested President of Iran | 9.74 office contested Governor of California |
| **V** Late - arriving evidence included a letter dated 17 December 1992 from William F. Ruddiman, who had become President of Iran after winning the 1980 Iranian presidential election . | **R** 2003 California gubernatorial recall election Arnold Schwarzenegger |
| **L** 7.05 description American paleoecologist | 11.99 description musical keyboard instrument |
| 6.72 occupation geologist | 8.24 Commons category Organs (music) |
| 6.71 instance of Human | 7.96 described by source Catholic Encyclopedia |
| 5.76 award received Lyell Medal | 5.19 obtained range of 1.71 to 1.80 |
| **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 . | **R** Baldwin was a strong backer of David Davis in the 2005 Conservative leadership election, having also supported him in the 2001 Conservative Party (UK) leadership election, having also supported him in the 2001 Conservative Party (UK) leadership election, having also supported him in the 2001 Conservative Party (UK) leadership election |
| **L** Allard studied clarinet under André Siewert of the Boston Symphony and saxophone under Lyle Bowen . | 6.04 description Jacobsite noble |
| **V** 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 . | 14.60 occupation musician |
| **R** 16.37 description American musician | 14.60 occupation musician |
| 15.44 instance of Human | 15.44 instance of Human |
| 14.06 occupation musician | 14.06 occupation musician |
| 13.60 sex or gender male | 13.60 sex or gender male |
| **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 . | 11.25 description Special election for the governorship of the U... |
| **R** 16.72 description German athlet | 16.72 description German athlet |
| 11.99 description musical keyboard instrument | 11.99 description musical keyboard instrument |
| 8.24 Commons category Organs (music) | 8.24 Commons category Organs (music) |
| 8.18 has part or parts organ case | 8.18 has part or parts organ case |
| **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 . | 11.02 instance of Human |
| **R** 11.72 description German athlet | 11.72 description German athlet |
| 13.01 description any unspecified or undetermined member of the ... | 13.01 description any unspecified or undetermined member of the ... |
| 8.96 Commons category Clarinets | 8.96 Commons category Clarinets |
| 8.85 award received Instrument of the Year | 8.85 award received Instrument of the Year |
| 8.66 described by source Armenian Soviet Encyclopedia | 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 |
|---|---|
| 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. |
| R | | |
| 1976 Austrian Grand Prix | John Watson |
| 9.40 instance of | Austrian Grand Prix |
| 9.28 description | 275th Formula 1 Championship Grand Prix |
| 8.16 label | 1976 Austrian Grand Prix |
| 8.13 point in time | time +1976-08-15T00:00:00Z timezone 0 before 0 ... |
| The 2007 Championship was won by John Higgins, who beat qualifier Mark Selby 18–13 in the final. |
| R | | |
| 2007 World Snooker Championship | John Higgins |
| 8.69 instance of | snooker tournament |
| 8.58 description | 2007 World Snooker Championship |
| 7.54 label | 2007 World Snooker Championship |
| 7.52 sponsor | 888 Holdings |
| Winged Foot member Uldis Balodis won three major titles: the 1927 U.S. Open, 1930 PGA Championship, and the 1931 British Open. |
| R | | |
| 1930 PGA Championship | Uldis Balodis |
| 12.42 instance of | PGA Championship |
| 12.26 description | golf tournament held in 1930 |
| 10.80 label | 1930 PGA Championship |
| 10.76 coordinate location | altitude None ... |
| Hotel du Lac is a 1984 Booker Prize-winning novel by English writer Lynn Paula. |
| R | | |
| Booker Prize | Lynn Paula |
| 8.06 instance of | literary award |
| 7.96 description | literary award |
| 7.19 Commons category | Man Booker Prize |
| 7.00 label | Booker Prize |
| 7.00 sex or gender | female |
| The reaction is named for Nobel Prize-winning chemist Georg Wittig. |
| R | | |
| Nobel Prize in Chemistry | Georg Wittig |
| 5.54 description | German chemist (1979 Nobel Prize) |
| 4.59 instance of | human |
| 4.45 occupation | chemist |
| 4.45 Commons category | Georg Wittig |
| 4.44 topic’s main category | Category:Nobel Prize in Chemistry |
| At the Golden Raspberry Awards, the film was nominated for Worst Actress (Miley Cyrus) and Worst Supporting Actor (Tyrell Lynch). |
| R | | |
| Golden Raspberry Award for Worst Supporting Actor | Tyrell Lynch |
| 10.40 description | college basketball player (2009–2009 Massachu ... |
| 8.68 instance of | human |
| 8.41 occupation | basketball player |
| 8.28 sex or gender | male |
| 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. |
| R | | |
| Academy Award for Best Production Design | Elia Meschak |
| 7.87 description | Congolese association football player |
| 6.55 instance of | human |
| 6.35 occupation | association football player |
| 6.34 Commons category | Meschak Elia |
| 6.34 Commons category | United States of America |
| 6.34 Commons category | Academy Award winners |
| 6.34 Commons category | Category:Best Art Direction |
| 6.34 Commons category | Academy Award for Best Production Design |

9251
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. |