Hyperbolic Disentangled Representation for Fine-Grained Aspect Extraction

Chang-You Tai[1], **Ming-Yao Li**[1], **Lun-Wei Ku**[1]
[1]Academia Sinica, Taipei, Taiwan
johnnyjana730@gmail.com, xuauul@gmail.com, lwku@iis.sinica.edu.tw

Abstract

Automatic identification of salient aspects from user reviews is especially useful for opinion analysis. There has been significant progress in utilizing weakly supervised approaches, which require only a small set of seed words for training aspect classifiers. However, there is always room for improvement. First, no weakly supervised approaches fully utilize latent hierarchies between words. Second, each seed word's representation should have different latent semantics and be distinct when it represents a different aspect. In this paper we propose HDAE, a hyperbolic disentangled aspect extractor in which a hyperbolic aspect classifier captures words' latent hierarchies, and an aspect-disentangled representation models the distinct latent semantics of each seed word. Compared to previous baselines, HDAE achieves average F1 performance gains of 18.2% and 24.1% on Amazon product review and restaurant review datasets, respectively. In addition, the embedding visualization experience demonstrates that HDAE is a more effective approach to leveraging seed words. An ablation study and a case study further attest the effectiveness of the proposed components.

Introduction

Researchers have begun to focus on aspect extraction, the automatic detection of fine-grained segments with predefined aspects (Hu and Liu 2004; Liu 2012; Pontiki et al. 2016), due to its potential for downstream tasks. For example, aspect extraction benefits users and customers when searching through review segments for aspects of interest on the Internet. Aspect extraction is also crucial for document summarization (Angelidis and Lapata 2018), recommendation justification (Ni, Li, and McAuley 2019), and review-based recommendation (Chin et al. 2018).

Aspect extraction research can be divided into supervised approaches, unsupervised approaches, and weakly supervised approaches.[1] Among these, many studies have been conducted on weakly supervised approaches (Karamanolakis et al. 2019; Angelidis and Lapata 2018; Zhuang et al. 2020) since they allow the model to be trained without substantial human-labeled data. For example, Angelidis and Lapata (2018) initialize fine-grained aspect representations using

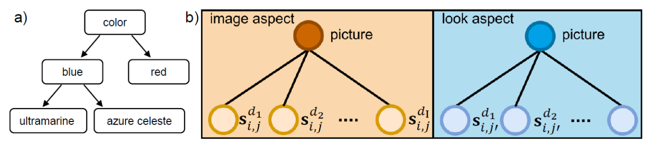

Figure 1: a) Seed word *color* and its hypernym pairs. b) An illustration of latent semantics under seed word *picture*. For example, in the TV domain's *image* aspect, *pixel of picture* $\mathbf{s}_{i,j}^{d_1}$ and *screen picture* $\mathbf{s}_{i,j}^{d_2}$ exist, wherease in the boot domain's *look* aspect, *cute picture* $\mathbf{s}_{i,j'}^{d_1}$ and *attractive picture* $\mathbf{s}_{i,j'}^{d_2}$ exist.

only a small number of descriptive keywords, or seed words, to identify highly salient opinions in review segments. Also, Karamanolakis et al. (2019) propose a student-teacher framework that more effectively leverages seed words by using a bag-of-words classifier teacher.

However, there is room for improvement in such seed word based methods. First, they neglect to consider the latent hierarchies between words, and it is assumed that capturing latent hierarchies between words will further improve seed word based methods on aspect inference, for instance by better identifying and organizing seed words and their hypernym pairs (Huang et al. 2020; López, Heinzerling, and Strube 2019). For example, as shown in Fig. 1(a), the general seed word *color* near the top can be used to find the more specific words *blue* or *green* in the middle, after which even more specific words can be found such as *ultramarine* or *azure celeste*. If seed words or their hypernym pairs exist in one review segment, the model can infer that it is of the corresponding aspect.

To allow the model to fully capture latent hierarchies between words, we introduce hyperbolic space (Nickel and Kiela 2017; Murty et al. 2018; Xu and Barbosa 2018; López, Heinzerling, and Strube 2019; López and Strube 2020). Compared to Euclidean space, hyperbolic space effectively encodes hierarchical structure information (Nickel and Kiela 2017), the latent hierarchies between words in this paper. In particular, when embedding tree-like structures, compared to the volume in Euclidean space, which leads to high distortion embeddings (Sa et al. 2018; Sarkar 2011), volume in hyperbolic space grows exponentially and can embed trees with arbitrarily low distortion (Sarkar 2011; Nickel and Kiela 2017). By virtue of such a hierarchy, a seed word based model

[1]We touch on unsupervised and supervised approaches in the related work section.

can better identify and utilize seed words and their hypernym words and thus achieve better aspect inference in hyperbolic space.

Second, existing seed word-based approaches model each seed word representation in a uniform manner while neglecting the fact that each seed word should have different latent semantics when conducting aspect extraction. For example, for the Amazon product review dataset (Angelidis and Lapata 2018), in the TV domain's *picture* aspect, the latent semantics under the seed word *picture* can be *pixel of the picture*, *screen picture*, or *HD picture*, as shown in Fig. 1(b). It is essential to select the most relevant latent semantics of the seed word when using the seed word *picture* to infer review aspects of segments. Furthermore, as shown in Fig. 1(b), the latent semantics of the seed word should be different in different aspects: this is also neglected by the current uniform representation. Such a uniform approach to modeling seed words tends to result in sub-optimal representations.

Thus, we propose HDAE, a hyperbolic disentangled aspect extractor which captures words' latent hierarchies and disentangles the latent semantics of each seed word. First, we propose a hyperbolic aspect classifier, using a hyperbolic distance function to calculate the relationship between the segment vector and the aspect representation generated from the seed word. Second, we introduce an aspect disentanglement module to model each seed word's latent semantics and then generate an aspect-refined representation of each review segment by selecting the most relevant latent semantics. In addition, we propose aspect-aware regularization to model each latent semantic meaning under its aspect scope while encouraging the independence of different latent semantic meanings. We conduct experiments on two datasets, demonstrating that HDAE achieves better aspect inference, which is further substantiated by embedding visualizations. We also provide two case studies to investigate HDAE's aspect inference ability compared with baselines without fully capturing words' latent hierarchies and the interpretability of the seed words' disentangled latent semantics.

We summarize our contributions: first, we propose a novel hyperbolic disentangled aspect extractor. To the best of our knowledge, this is the first work to investigate how to leverage hyperbolic components and disentangled representations for weakly supervised approaches to aspect extraction.[2] Second, we propose a hyperbolic aspect classifier which captures word's latent hierarchies and generates associations between the review segment and aspects of interest. Third, we introduce the aspect disentanglement module and aspect-aware latent semantic regularization to model the latent semantic meaning of each seed word. Experiments and a case study demonstrate the effect of the proposed methods for aspect extraction.

Related Work

Aspect Extraction In addition to weakly supervised approaches, there are also supervised approaches and unsupervised approaches. Supervised neural networks achieve better performance than traditional rule-based approaches by viewing aspect extraction as a sequence labeling problem which can be tackled with hidden Markov models (Jin, Ho, and Srihari 2009), conditional random fields (Yang and Cardie 2012; Mitchell et al. 2013), or recurrent neural networks (Wang et al. 2016; Liu, Joty, and Meng 2015). However, supervised approaches require large amounts of labeled data for training. Unsupervised approaches, in contrast, do not use annotated data. Early examples are latent Dirichlet allocation (LDA)-based methods (Chen, Mukherjee, and Liu 2014; García-Pablos, Cuadros, and Rigau 2018; Shi et al. 2018). Recently, neural network-based methods (Iyyer et al. 2016; Srivastava and Sutton 2017; He et al. 2017; Luo et al. 2019; Shi et al. 2020) have shown remarkable performance and have outperformed LDA-based methods. However, unsupervised approaches are not effective when used directly for aspect extraction (Karamanolakis et al. 2019; Angelidis and Lapata 2018; Tulkens and van Cranenburgh 2020). For example, many-to-one mapping or high-resolution selective mapping is required by He et al. (2017) and Shi et al. (2020) to manually associate the model-inferred aspect with gold-standard aspects.

Hyperbolic representations have been used to model complex networks (Krioukov et al. 2010; Nickel and Kiela 2017, 2018; Tay, Luu, and Hui 2017; Gülçehre et al. 2018; Huang et al. 2020) and have proven more suitable than Euclidean space in representing hierarchical data (Sala et al. 2018; Nickel and Kiela 2017). For example, López and Strube (2020) introduce hyperbolic representations to capture latent hierarchies arising from the class distribution for multi-class multi-label classification. Aly et al. (2019) use Poincaré embeddings to improve existing methods for domain-specific taxonomy induction. Le et al. (2019) propose utilizing hyperbolic representations to infer missing hypernymy relations. Sun et al. (2021) show that points in hyperbolic space can be more concentrated while maintaining the desired separation and revealing nuanced differences. To our knowledge, this is the first work to apply hyperbolic representations to weakly supervised approaches for aspect extraction.

Disentangled representations improve model performance by identifying and disentangling latent explanatory factors in the observed data (Yoshua Bengio and Vincent 2012) and have shown their success in the NLP domain (Shen et al. 2017; Zhao et al. 2018; Chen et al. 2019; Hu et al. 2017). For instance, Hu et al. (2017) propose disentangled representations with designated semantic structure, which generates sentences with dynamically specified attributes. Hou et al. (2021) derive disentangled representations which separate the distinct and informative factors of variations to improve content-based detection. Disentangled representation has been successively applied to the recommendation (Ma et al. 2019b,a; Hu et al. 2020) and computer vision (Liu et al. 2020; Dupont 2018) domains. For example, Wang et al. (2020) model diverse relationships and disentangle user intents to achieve better-performing representations. To our knowledge, this is the first work to apply disentangled representations to weakly supervised approaches for aspect extraction.

Preliminaries

Problem formulation The goal of aspect extraction is to predict an aspect category $a_i \in A_{\mathrm{C}} = \{a_j\}_{j=1}^{\mathrm{K}}$ given a

[2]The codes is at https://github.com/johnnyjana730/HDAE/

review segment (e.g., sentence, clause) $x^s = \{x_1, x_2, ..., x_{\mathrm{T}}\}$ from a specific domain d_{C} (e.g., laptop bags, TVs), where the review segments are created by splitting each review in the corpus; x_i is the word index in the segment; a_i is an aspect and A_{C} refers to the aspect set pertaining to domain d_{C}; K is the number of total aspects and T is the segment's length. For every aspect $a_i \in A_{\mathrm{C}}$, a small number of seed words $[s_{i,1}, s_{i,2}, ..., s_{i,N}]$ are provided during training. The classifier predicts K aspect probabilities $\mathrm{p}_s^a = \langle \mathrm{p}_s^{a_1}, ..., \mathrm{p}_s^{a_{\mathrm{K}}} \rangle$ given a test review segment x^s and the seed words.

Hyperbolic Geometry We introduce two hyperbolic models:[3] the Poincaré ball model and the Klein model.

The Poincaré ball model is defined as a Riemannian manifold $\mathcal{P}^n = (\beta, g_{\mathbf{x}}^{\beta})$, where $\beta^n = \{\mathbf{x} \in \mathbb{R}^n : ||\mathbf{x}|| < 1\}$ is an open unit ball, with the metric tensor $g_{\mathbf{x}}^{\beta} = \lambda_{\mathbf{x}}^2 g^E$, where $\lambda_{\mathbf{x}} = \frac{2}{1-||\mathbf{x}||^2}$; g^E is the Euclidean metric tensor. The distance on the manifold is defined as

$$d_{\mathcal{P}}(\mathbf{x}, \mathbf{y}) = \operatorname{arcosh}\left(1 + 2\frac{||\mathbf{x}-\mathbf{y}||^2}{(1-||\mathbf{x}||^2)(1-||\mathbf{y}||^2)}\right). \quad (1)$$

The Klein model is given by $\mathcal{K}^n = \{\mathbf{x} \in \mathbb{R}^n : ||\mathbf{x}|| < 1\}$ and is often used for aggregation since the Einstein midpoint (Gülçehre et al. 2018) can be easily computed in the Klein model. Formally, a point in the Klein model can be obtained from Poincaré coordinates by

$$\mathcal{P}^n \to \mathcal{K}^n : \pi_{\mathcal{P}\to\mathcal{K}}(\mathbf{x}_{\mathcal{K}}) = \frac{2\mathbf{x}_{\mathcal{P}}}{1 + ||\mathbf{x}_{\mathcal{P}}||^2} \quad (2)$$

and the backward transition formulas

$$\mathcal{K}^n \to \mathcal{P}^n : \pi_{\mathcal{K}\to\mathcal{P}}(\mathbf{x}_{\mathcal{P}}) = \frac{\mathbf{x}_{\mathcal{K}}}{1 + \sqrt{1 - ||\mathbf{x}_{\mathcal{K}}||^2}}. \quad (3)$$

For the Poincaré ball model, the exponential map, from tangent space to hyperboloid manifold, $\exp_{\mathbf{x}}$: $\mathcal{T}_{\mathbf{x}}\mathcal{P} \to \mathcal{P}$, and the logarithmic map, from hyperboloid manifold to tangent space, $\log_{\mathbf{x}}$: $\mathcal{P} \to \mathcal{T}_{\mathbf{x}}\mathcal{P}$, can be found in Liu, Nickel, and Kiela (2019). For simplicity, we denote $d_{\mathcal{P}}^{\exp}$ as the hyperbolic distance of the tangent space vector after applying the exponential map:

$$d_{\mathcal{P}}^{\exp}(\mathbf{x}, \mathbf{y}) = d_{\mathcal{P}}(\exp_0(\mathbf{x}), \exp_0(\mathbf{y})). \quad (4)$$

Methodology

Euclidean Aspect Extractor

Our work builds on the seed word based model developed by Angelidis and Lapata. We describe the method, including segment representation generation and the aspect classifier.

Segment Representation For each review segment $x^s = \{x_1, x_2, ..., x_{\mathrm{T}}\}$, the segment representation $\mathbf{v}_s$ is generated by a weighted sum of an individual word:

$$\mathbf{v}_s = \sum_{i=1}^{n} c_i \mathbf{v}_{x_i} \quad (5)$$

$$c_i = \frac{\exp(u_i)}{\sum_{j=1}^{n} \exp(u_j)}; u_i = \mathbf{v}_{x_i}^{\top} \cdot \mathbf{M} \cdot \mathbf{v}_s', \quad (6)$$

where $\mathbf{v}_{x_i}$ is the vector of the i-th word x_i; $\mathbf{v}_s'$ is average of the segment's word vector; and $\mathbf{M} \in \mathbb{R}^{d\times d}$ denotes the attention matrix.

[3]For more details; see Robbin and Salamon (2011).

Euclidean Aspect Classifier To predict a probability distribution over K aspects, the vector $\mathbf{v}_s$ is fed to a hidden classification layer followed by the softmax function:

$$\mathrm{p}_s^a = \operatorname{softmax}(\mathbf{W}\mathbf{v}_s + \mathbf{b}), \quad (7)$$

where $\mathbf{W}$ and $\mathbf{b}$ are trainable parameters. To focus on the aspect of interest, for each aspect a_i, which has seed words $[s_{i,1}, s_{i,2}, ..., s_{i,N}]$, the model generates the aspect vector $\mathbf{a}_i$ by using the labeled aspect seed words:

$$\mathbf{a}_i = \sum_{j}^{N} \mathbf{z}_{i,j}\mathbf{s}_{i,j}; \mathbf{A} = [\mathbf{a}_i^{\top}; ...; \mathbf{a}_{\mathrm{K}}^{\top}], \quad (8)$$

where $\mathbf{A} \in \mathbb{R}^{\mathrm{K}\times d}$ denotes the aspect matrix; and $\mathbf{s}_{i,j}$ denotes the j-th seed word representation; the weight vectors $\mathbf{z}_{i,j}$ are determined by the method mentioned in Angelidis and Lapata (2018); and N is the number of seed words. Then, the segment reconstructed vector $\mathbf{r}_s$ is generated based on the aspect vector:

$$\mathbf{r}_s = \mathbf{A}^{\top}\mathrm{p}_s^{\mathrm{asp}}. \quad (9)$$

To optimize the performance, the model is trained by reconstruction loss, which maximizes the distance between inner product $\mathbf{r}_s\mathbf{v}_s$ and $\mathbf{r}_s\mathbf{v}_n$, where $\mathbf{v}_{n_i}$ is the vector of a randomly sampled negative segment.

$$J_r(\theta) = \sum_{x^s\in C}\sum_{i=1}^{k_n} \max(0, 1 - \mathbf{r}_s\mathbf{v}_s + \mathbf{r}_s\mathbf{v}_{n_i}), \quad (10)$$

Hyperbolic Disentangled Aspect Extractor

Here, we present HDAE, a hyperbolic aspect classifier with an aspect disentanglement module proposed to model multiple latent semantic meanings for each seed word according to its aspect category.

Hyperbolic Aspect Classifier To infer the review segment's aspect probability $\mathrm{p}_s^{a_i}$ in hyperbolic space, we follow Balazevic, Allen, and Hospedales (2019) in using the hyperbolic distance function and biases to calculate the relationship between segment vector $\mathbf{v}_s$ and aspect representation $\mathbf{a}_i$ as

$$\mathrm{p}_s^{a_i} = -d_{\mathcal{P}}^{\exp}(\mathbf{v}_s, \mathbf{a}_i)^2 + b_v + b_{a_i}. \quad (11)$$

Then, to generate the reconstructed embedding $\mathbf{r}_s$, the Einstein midpoint is used to aggregate hyperbolic aspect weights, with a simple form in the Klein disk model:

$$\mathbf{r}_s = \log_0(\pi_{\mathcal{K}\to\mathcal{P}}(\sum_{a_i\in A_{\mathrm{C}}} \frac{k_i\gamma(\mathbf{a}_i^{\mathcal{K}})}{\sum_j k_j\gamma(\mathbf{a}_j^{\mathcal{K}})}\mathbf{a}_i^{\mathcal{K}})) \quad (12)$$

$$k_i = \exp(\beta \mathrm{p}_s^{a_i} - c), \quad (13)$$

where $\mathbf{a}_i^{\mathcal{K}} = \pi_{\mathcal{P}\to\mathcal{K}}(\mathbf{a}_i^{\mathcal{P}})$; $\mathbf{a}_i^{\mathcal{P}}$ denotes the Poincaré aspect embedding; $\mathbf{a}_i^{\mathcal{P}} = \exp_0(\mathbf{a}_i)$; β and c are set parameters; and Lorentz factors $\gamma(t) = \frac{1}{(1-||t||^2)^{1/2}}$.

Aspect Disentanglement Module To generate multiple latent semantic meanings for each seed word, we propose a disentangled semantic representation. Then, we present aspect-aware regularization, which models latent semantic vectors for each seed word, after which we discuss refined seed word representation.

Disentangled Semantic Representation For aspect a_i, we devise a representation function to output a disentangled semantic vector $\mathbf{s}_{i,j}^{d}$ for the j-th seed word $s_{i,j}$, which is composed of I independent components:

$$\mathbf{s}_{i,j}^{d} = (\mathbf{s}_{i,j}^{d_1}, \mathbf{s}_{i,j}^{d_2}, \mathbf{s}_{i,j}^{d_3}, ..., \mathbf{s}_{i,j}^{d_{\mathrm{I}}}), \tag{14}$$

where disentangled semantic vector $\mathbf{s}_{i,j}^{d_k}$ is generated by adding a standard Gaussian random variable to the original seed word representation $\mathbf{s}_{i,j}$.

Aspect-Aware Regularization This models the latent semantic representation of each seed word according to its aspect category and has three objectives, as shown in Fig. 2: (a) seed word dependence, (b) latent semantic independence, and (c) aspect scope confinement, which are controlled by latent semantic modeling distances d_1, d_2, and d_3.

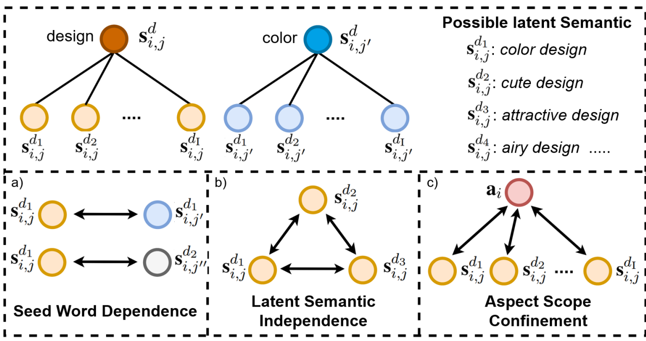

Figure 2: The proposed aspect disentanglement module generates disentangled semantic representations for each seed word and models latent semantics using (a) seed word dependence, (b) latent semantic independence, and (c) aspect scope confinement.

Seed Word Dependence The interdependence between seed word pairs sheds light on the modeling of the seed word's latent semantics within the scope of its aspect. For example, for seed word *design* in the boot domain's *look* aspect, the latent semantic meaning, which facilitates fine-grained aspect inference, can be *color design*, *design style*, *cute design*, and *attractive design*. The desired latent semantic meaning can be modeled by narrowing the gap between either the latent semantic meaning of *design* and the latent semantic meanings of other seed words, such as *color*, *style*, *cute*, and *attractive* in the same *look* aspect. Likewise, in the TV domain's *service* aspect, latent semantic meanings *shipping service*, *replacement service*, and *delivery service* can be generated by minimizing the distance between either the latent semantic meaning of *service* and that of *shipping*, *replacement*, and *delivery*, which are seed words in the same aspect.

To model the interdependence of seed word pairs, we use the hyperbolic distance function $d_{\mathcal{P}}()$ to achieve fine-grained relationship modeling, since hyperbolic space offers the ability to not only preserve hierarchical (tree-like) information (Nickel and Kiela 2017; Zhang and Gao 2020; Gülçehre et al. 2018; Chami et al. 2019) but also nuanced differences (to better group them) (Sun et al. 2021; Tai et al. 2021) and outperforms Euclidean counterparts in various kinds of data (Zhang and Gao 2020; Gülçehre et al. 2018; Chami et al. 2019, 2020; Sun et al. 2021; Tai et al. 2021). Thus, it is assumed that with more space (hyperbolic space) to organize points, the model can divide disentangled representations and better group them. Given seed word pairs such as *design* $s_{i,j}$ and *color* $s_{i,j'}$ in the specific aspect, we require at least one latent semantic pair distance to be close enough:

$$\begin{aligned} \mathrm{sim}(s_{i,j}, s_{i,j'}) = \operatorname{argmin}\{ \\ d_{\mathcal{P}}^{\exp}(\mathbf{s}_{i,j}^{d_k}, \mathbf{s}_{i,j'}^{d_{k'}}) | \mathbf{s}_{i,j}^{d_k} \in \mathbf{s}_{i,j}^{d}, \mathbf{s}_{i,j'}^{d_{k'}} \in \mathbf{s}_{i,j'}^{d} \} \end{aligned} \tag{15}$$

$$\begin{aligned} J_{d_1}(\theta) = \sum_{a_i \in A_C} \sum_{j=1}^{N} \sum_{j'=j+1}^{N} \max(0, \\ \mathrm{sim}(s_{i,j}, s_{i,j'}) - d_1), \end{aligned} \tag{16}$$

where sim outputs the minimal distance from all possible seed word latent semantic meaning pairs; d_1 is the inter seed word alignment distance, which maintains two latent semantic meanings within a certain distance. Intuitively, for different aspect word pairs, the alignment score should be different, as in Wang et al. (2020). For example, in the boot domain's *look* aspect, the seed word dependence between *design* and *color* should be more significant than *design* and *going*. We leave this to future work.

Latent Semantic Independence Latent semantic meanings should be distinct from each other. Independent latent semantic meanings reduce redundancy and confusion in aspect inference. To achieve this, we maintain the distance between the seed word's latent semantic meanings.

$$\begin{aligned} J_{d_2}(\theta) = \sum_{a_i \in A_C} \sum_{j}^{N} \sum_{k=1}^{\mathrm{I}} \sum_{k'=k+1}^{\mathrm{I}} \max \\ (0, d_2 - d_{\mathcal{P}}^{\exp}(\mathbf{s}_{i,j}^{d_k}, \mathbf{s}_{i,j}^{d_{k'}})), \end{aligned} \tag{17}$$

where d_2 is the latent semantic distance.

Aspect Scope Confinement For each seed word, all latent semantic meanings should be limited in terms of aspect scope. For example, in the boot domain's *color* aspect, all latent semantic meanings of seed word *style* should refer to *color's style*. However, in the *look* aspect, all latent semantic meanings of the same seed word *style* should refer to *outlook style*. To thus limit all latent meanings of a seed word to its aspect scope, we introduce another regularization:

$$J_{d_3}(\theta) = \sum_{a_i \in A_C} \sum_{j}^{N} \sum_{k=1}^{\mathrm{I}} \max(0, d_{\mathcal{P}}^{\exp}(\mathbf{s}_{i,j}^{d_k}, \mathbf{a}_i) - d_3), \tag{18}$$

where d_3 is the aspect scope confinement distance and $\mathbf{a}_i$ is the aspect representation from Eq. 8. Note compared to seed word dependence and Eq. 16, which focuses on dependence between seed word pairs, Eq. 18 ensures all latent semantic meanings are modeled within the specific aspect scope.

Refined Seed Word Representation This constructs refined seed representations based on its latent semantics. For each seed word, the latent semantics should be independent from each other; only one latent semantic meaning should be used to find the aspect relevant content. For example, for the boot domain's *look* aspect, possible latent semantics of seed word *style* include *cute style*, *casual style*, or *attractive style*; as we can see these latent semantics are of different meanings, and combining them together may led to a sub-optimal seed word representation. Also, according to each review segment, the most relevant latent semantic meaning should be selected when predicting its aspect distribution. Thus, we introduce the Gumbel softmax (Jang, Gu, and Poole 2017), a differentiable softmax function for generating discrete variables, and use this to generate the desired refined seed word representation $\mathbf{s}^r_{i,j}$ according to each segment $\mathbf{v}_s$:

$$\mathbf{s}^r_{i,j} = \sum_{k=1}^{\mathrm{I}} g_k \mathbf{s}^{d_k}_{i,j},\ g_k = \frac{c_k}{\sum_{k'} c_{k'}}, \tag{19}$$

$$c_k = \exp\left(\frac{-d_{\mathcal{P}}(\mathbf{v}_s, \mathbf{s}^{d_k}_{i,j})}{\tau}\right) \tag{20}$$

where τ, the temperature parameter, controls the extent to which the output becomes a one-hot vector. With the refined seed word representation $\mathbf{s}^r_{i,j}$, the aspect representation can be generated by Eq. 8.

Algorithm 1: HDAE Learning

Input: review segments $S = \{s \mid s \in \mathrm{C}\}$, aspect seed words
Initialize HDAE parameter with pre-trained word vector
foreach *epoch* **do**
  for $x^s \in S$ **do**
    Generate segment embedding $\mathbf{v}_s$ (Eq 5)
    for $i \leftarrow 1$ **to** K **do**
      Generate refined aspect seed word vector $\mathbf{s}^r_{i,j}$ (Eq 19)
      Calculate aspect embedding $\mathbf{a}_i$ (Eq 8)
      Generate aspect probability $\mathrm{p}^{a_i}_s$ (Eq 11)
    Generate reconstructed embedding $\mathbf{r}_s$ (Eq 12)
    Calculate objective J (Eq 21)
  Update parameters by Adam optimizer

Learning Algorithm

The formal description of the above aspect inference process is presented in Algorithm 1. To train HDAE, we rely on the previously introduced reconstruction loss J_r (Eq. 10). Since the reconstruction objective only provides a weak training signal (Angelidis and Lapata 2018), the distillation objective J_d from the teacher (Karamanolakis et al. 2019) is used to provide an additional training signal. Also, the disentangled modeling objectives J_{d_1}, J_{d_2}, and J_{d_3} are used to model each latent semantic meaning according to its aspect category. Thus, the overall objective is

$$J(\theta) = J_r(\theta) + \lambda J_d(\theta) + J_{d_1}(\theta) + J_{d_2}(\theta) + J_{d_3}(\theta). \tag{21}$$

The λ controls the influence of the distillation objective loss.

Experiments and Results

Datasets We used Amazon product reviews from the OPOSUM dataset (Angelidis and Lapata 2018) and restaurant reviews from the SemEval-2016 Aspect-based Sentiment Analysis task (Pontiki et al. 2016). The Amazon product reviews cover six domains, ranging from laptop bags (Bags), bluetooth headsets (B/T), boots, keyboards (KBs), and televisions (TVs) to vacuums (VCs). The restaurant reviews dataset covers six languages: English (En), Spanish (Sp), French (Fr), Russian (Ru), Dutch (Du), and Turkish (Tur). During training, seed words are provided but not segment aspect labels. Details are provided in the appendix.

	Product review domain					
Model	Bags	KBs	Boots	B/T	TVs	VCs
LDA-Anchors	33.5	34.7	31.7	38.4	29.8	30.1
ABAE	38.1	38.6	35.2	37.6	39.5	38.1
SSCL	61.0	60.6	57.3	65.2	64.6	57.2
SSCL-BT	65.5	62.3	60.4	69.5	67.0	61.0
SSCL-BT*	56.5	61.7	41.5	51.4	58.2	52.4
MATE	46.2	43.5	45.6	52.2	48.8	42.3
MATE-MT	48.6	45.3	46.4	54.5	51.8	47.7
TS-Teacher	59.3	58.2	50.6	63.3	61.0	58.4
TS-ATT	58.7	57.0	52.6	67.6	63.2	58.8
TS-BT	59.1	59.0	53.9	65.8	66.1	61.0
HDAE	**68.8**	**72.2**	**64.0**	**72.0**	**71.2**	**66.9**

Table 1: Micro-averaged F1 for 9-class EDU-level aspect detection in product reviews

Baseline LDA-Anchors (Lund et al. 2017), an interactive topic model which utilizes seed words as "anchors" to identify the segment aspect. **ABAE** (He and Chua 2017), an unsupervised method which adopts reconstruction loss to make the reconstructed embedding similar to a segment vector. This requires a manual mapping between the model-inferred aspect and gold-standard aspects. **SSCL** (Shi et al. 2020), an unsupervised method that uses a contrastive learning algorithm and knowledge distillation for aspect inference. For manual mapping, the high-resolution selective mapping (HRSMap) is used. **MATE*** (Angelidis and Lapata 2018), a seed-based weakly supervised method which generates predefined aspect representations by seed word vector. This can be trained by an extra multitask training objective (MT).[4] **TS-*** (Karamanolakis et al. 2019), a seed based weakly supervised method which adopts a teacher-student iterative co-training framework, where the teacher (TS-Teacher) is a bag-of-words classifier based on seed words and the student uses the attention-weighted average of word2vec embeddings (TS-ATT). **Gold-***, supervised models trained using ground truth aspect labels, only available for restaurant reviews, and not directly comparable with other weakly supervised baselines (Karamanolakis et al. 2019).

Note that for SSCL and TS, the BERT model also can be used as the encoder (SSCL-BT, TS-BT). The results of the compared models are obtained from the corresponding published papers. We also report our re-implemented version of SSCL-BT*. We do not provide the ABAE and SSCL results for restaurant reviews for non-English datasets, since this requires domain knowledge for manual aspect mapping.[5]

Implementation Details For HDAE and other models, detailed hyper-parameter settings are given in the appendix.

[4] MT cannot be applied in restaurant reviews since it requires datasets from different domains but the same language.

[5] We report ABAE and SSCL results for EN restaurant reviews in the appendix in our arxiv version.

Model	Restaurant review domain En	Sp	Fr	Ru	Du	Tur
W2V-Gold	58.8	50.4	50.4	69.3	51.4	55.7
BERT-Gold	63.1	51.6	50.6	64.6	53.5	55.3
HDAE-Gold	**70.5**	**72.5**	**65.4**	**67.9**	**73.8**	**65.4**
LDA-Anchors	28.5	17.7	13.1	14.8	25.9	27.7
MATE	41.0	24.9	25.8	18.4	36.1	39.0
MATE-UW	40.3	18.3	27.8	21.8	31.5	25.2
TS-Teacher	44.9	41.8	34.1	54.4	40.7	30.2
TS-ATT	47.8	41.7	32.4	59.0	42.1	42.3
TS-BT	51.8	42.0	39.2	58.0	43.0	45.0
HDAE	**57.9**	**65.7**	**48.6**	**62.9**	**57.2**	**50.8**

Table 2: Micro-averaged F1 for 12-class sentence-level aspect detection in restaurant reviews

Experimental Results

Overall Inference Performance Tables 1 and 2 show the results for aspect extraction on both datasets. We observe that HDAE achieves superior performance. For example, in Amazon product reviews, compared to TS-W2V, HDAE yields F1 performance gains of 16.0%, 8.1%, 31.9%, 24.7%, 11.3%, and 17.4% on Bags, KBs, Boots, B/T, TVs, and VCs, respectively; similar trends are observed in the restaurant review dataset. Moreover, the reduction in the parameter size of HDAE is also remarkable: 97.8% versus TS-BT.[6] These results demonstrate the effectiveness of the proposed hyperbolic disentangled aspect extractor (HDAE).

We also observe the weakly unsupervised approaches MATE* and TS-* significantly outperform the unsupervised approaches LDA-Anchors and ABAE, suggesting the effectiveness of seed words. Note our reproduced SSCL-BT* does not consistently outperform MATE, perhaps because SSCL-BT relies heavily on the quality of initial k-means centroids since poorly initialized centroids may cause model-inferred aspects after training to lack good coverage for gold-standard aspects, and thus make manual mapping more difficult.

Compared to fully supervised (*-Gold) models, HDAE dramatically reduces the performance gap between weakly supervised approaches and fully supervised approaches and even outperforms in Spanish and Dutch restaurant reviews. Moreover, we investigate HDAE's performance when given ground-truth aspect labels denoted by HDAE-Gold in Table 2. The result shows that HDAE-Gold outperforms both W2V-Gold and BERT-Gold, showing the effectiveness of the proposed hyperbolic disentangled based approach. Note that we vary the ratio of ground-truth aspect labels and compare model performance for different label ratios; these result are provided in the appendix.

Inference Performance per Aspect Here we investigate the abilities that seed word based approaches infer on different aspects, shown in Fig. 3(a) and (b). First, we observe that TS-W2V puts a greater focus on the general aspect, possibly because the teacher always predicts review segments as general aspect if no seed word appears. Second, compared to MATE and TS-W2V, HDAE yields better inference performance in almost all aspects without putting excessive bias on certain predictions, showing better aspect inference ability.

[6]The parameter sizes of HDAE and TS-BT are 2.5M and 109.5M.

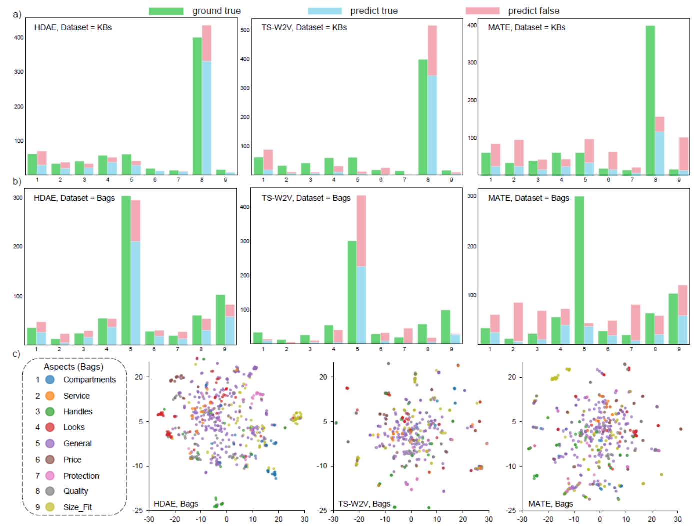

Figure 3: Inference performance per aspect of HDAE, TS-W2V, and MATE on the a) KBs and b) Bags datasets. On Bags, we use t-SNE to compare c) the embedding of each model, where the different colors represent different aspects.

Ablation	Bag	KBs	B/T	Boots	TV	VCs
HDAE	68.8	72.2	72.0	64.0	71.2	66.9
HDAE ($\lambda = 0$)	67.3	65.6	70.1	60.5	54.1	59.1
MATE	46.2	43.5	52.2	45.6	48.8	42.3

Table 3: HDAE ablation study. The λ is the ratio of distillation objective loss. When λ is 0, the distillation objective J_d is not used.

To further investigate the performance on Bags,[7] we compared sentence vectors $\mathbf{v}_s$ of each model[8] by using t-SNE to visualize vectors, as shown in Fig. 3(c). We find that a well-differentiated sentence vector benefits the model's aspect inference ability. First, compared to HDAE and MATE, less-separated vectors are found for TS-W2V, which correlates to the fact that TS-W2V performs poorly in every aspect except the general aspect. Second, we observe for both the *looks* (red) and *size fit* (yellow) aspects, differentiated vector clusters are found in HDAE and MATE, which correlates to the good accuracy in both aspects. Third, compared to MATE, the differentiated vector clusters of the *handles* (green), *protection* (pink), *price* (brown), and *compartments* (blue) aspects explain HDAE's better inference ability in those aspects.

Ablation Study and Parameter Sensitivity

To verify the effectiveness of the proposed components, we conducted an ablation study for HDAE, as shown in Table **??**. After removing the hyperbolic aspect classifier (3) and aspect disentangle module (4), we observe drops in performance, indicating the effect of the proposed components. Note that (4), which only contains the hyperbolic aspect classifier, outperforms the MATE-* and TS-* models, showing that the proposed hyperbolic aspect classifier effectively leverages the seed words. Furthermore, removing the hyperbolic distance function for the disentangled aspect representation, shown by (2), degrades performance, suggesting that the hyperbolic

[7]For other datasets, refer to the appendix in our arxiv version.

[8]For HDAE, the hyperbolic sentence vector $\exp_0(\mathbf{v}_s)$ is used.

a) The keyboard works very well. GT: General Seed Words: think, recommend, purchase, using, unit, star, microsoft		
HDAE: General ✓	MATE: General ✓	TS-W2V: General ✓
b) The color is nice, more light blue. GT: Color Seed Words: color, love, style, unbelievably, gorgeous, blue		
HDAE: Color ✓	MATE: Color ✓	TS-W2V: Color ✓
c) What I received is a grayish brown shoe. GT: Color Seed Words: color, love, style, unbelievably, gorgeous, blue		
HDAE: Color ✓	MATE: General ✗	TS-W2V: Price ✗
d) It's not leather. GT: Quality Seed Words: quality, material, handle, poor, broke, durable, month		
HDAE: Quality ✓	MATE: Handles ✗	TS-W2V: General ✗
e) On the other hand, find it to be too stiff. GT: Comfort Seed Words: feel, comfortable, mushy, key, like, difficult		
HDAE: Comfort ✓	MATE: Function ✗	TS-W2V: General ✗

Table 4: Comparison of predictions on sample Product review segments between HDAE, MATE, and MATE. For each review segment, the ground truth (GT) aspect and its corresponding seed words are provided.

distance function is indispensable for modeling the latent semantic meanings of seed words.

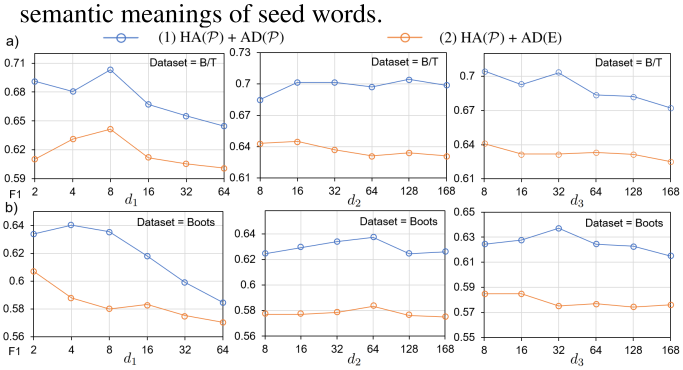

Figure 4: Micro-averaged F1 scores of (1) and (2) with different d_1, d_2, and d_3 on a) B/T and b) Boots datasets

Then, we investigated the sensitivity of latent semantic modeling distance d_1, d_2, and d_3[9] on (1) and (2), as shown in Fig. 4. We offer the following observations. First, both (1) and (2) achieve the best results when a small d_1, e.g, $d_1 \leq 8$, is set, demonstrating the importance of narrowing the gap between seed word pairs when modeling latent semantic meanings. Also, (1) and (2) both perform better when a large d_2, e.g, $d_2 \geq 64$, is set, verifying the importance of independence modeling. Last, (1) and (2) achieve the best performance when d_3 is set to around 8 to 32, perhaps due to the strong regularization on each latent semantic meaning introduced when d_3 is too large.

Case study

To more closely investigate the aspect inference ability of HDAE, we compare the predictions made by HDAE, MATE, and TS-W2V, the results of which are shown in Table 4. For the example in Table 4(b), we see that the review segment contains keywords such as *color* and *blue* which are explicitly captured in aspect seed words. All models correctly infer and review the segment's aspect. However, for cases in Table 4(c,d,e), the reviews' segments do not explicitly match

[9]For other parameters, please refer to the appendix in our arxiv version.

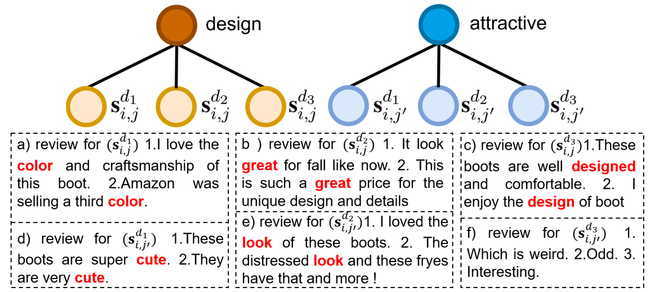

Figure 5: Interpretability of latent semantic meanigs of seed words. Best viewed in color.

their aspect seed words but instead match the hyponymic relations (is-a) present between seed words and review segments. For example, there are hierarchical relations such as *grayish* brown is a *color*, *leather* is a *material*, and *stiff* is a type of *difficult* for cases in Table 4(c,d,e). We find only HDAE correctly recognizes the review segments' aspects. We thus conclude HDAE captures and utilizes hyponymic relations (is-a) present between seed words and review segments, deriving reasonable aspect inference for each review segment and thus achieving better performance. Analogous behavior is observed for other cases in the appendix.

To explore the interpretability of the seed words' latent semantic meanings, we conducted a case study in which we randomly selected review segments from the boot domain's *look* aspect and investigated its association with each aspect of latent semantic meaning. Figure 5 shows the review segments captured by each seed word's latent semantic meaning: we find that each aspect's latent semantics focus on a distinct type of review segment. For example, for the seed word *design*, the latent semantic meaning $\mathbf{s}_{i,j}^{d_1}$ focuses on segments with *color* information, whereas $\mathbf{s}_{i,j}^{d_2}$ focuses on segments with the *great* keyword. Likewise, for the seed word *attractive*, the latent semantic meaning $\mathbf{s}_{i,j'}^{d_1}$ focuses on segments with *cute* information, whereas $\mathbf{s}_{i,j'}^{d_3}$ focuses on segments with *unattractive* information. These results demonstrate that the proposed aspect disentanglement module assists HDAE in modeling different latent semantics for each seed word. Also, HDAE finds the most relevant latent semantic meanings for each review segment, explaining the improvements in the aspect inference ability.

Conclusions and Future Work

We present HDAE, a hyperbolic disentangled aspect extractor which includes a hyperbolic aspect classifier and an aspect disentanglement module. On two datasets, HDAE, with its 97.8% reductions in parameter size versus TS-BT, shows superior aspect inference ability, further substantiated by an embedding visualization. The effect of the proposed components is proved by an ablation study, a parameter sensitivity study, and a case study.

In the future, we plan to explore the proposed module on other aspect-based sentiment analysis (ABSA) subtasks. We would also like to further improve the performance of the proposed components, for instance by setting up alignment scores for different aspect word pairs when modeling seed word dependence.

Acknowledgement

This research is partially supported by Ministry of Science and Technology, Taiwan under the project contract 110-2221-E-001-001- and 110-2634-F-002-051.

References

Aly, R.; Acharya, S.; Ossa, A.; Köhn, A.; Biemann, C.; and Panchenko, A. 2019. Every child should have parents: A taxonomy refinement algorithm based on hyperbolic term embeddings. *arXiv preprint arXiv:1906.02002*.

Angelidis, S.; and Lapata, M. 2018. Summarizing opinions: Aspect extraction meets sentiment prediction and they are both weakly supervised. *arXiv preprint arXiv:1808.08858*.

Balazevic, I.; Allen, C.; and Hospedales, T. M. 2019. Multi-relational Poincaré Graph Embeddings. *CoRR*, abs/1905.09791.

Chami, I.; Wolf, A.; Juan, D.; Sala, F.; Ravi, S.; and Ré, C. 2020. Low-Dimensional Hyperbolic Knowledge Graph Embeddings. *CoRR*, abs/2005.00545.

Chami, I.; Ying, R.; Ré, C.; and Leskovec, J. 2019. Hyperbolic Graph Convolutional Neural Networks. *CoRR*, abs/1910.12933.

Chen, M.; Tang, Q.; Livescu, K.; and Gimpel, K. 2019. Variational sequential labelers for semi-supervised learning. *arXiv preprint arXiv:1906.09535*.

Chen, Z.; Mukherjee, A.; and Liu, B. 2014. Aspect extraction with automated prior knowledge learning. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 347–358.

Chin, J. Y.; Zhao, K.; Joty, S.; and Cong, G. 2018. ANR: Aspect-based neural recommender. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 147–156.

Dupont, E. 2018. Learning Disentangled Joint Continuous and Discrete Representations. arXiv:1804.00104.

García-Pablos, A.; Cuadros, M.; and Rigau, G. 2018. W2VLDA: Almost unsupervised system for aspect based sentiment analysis. *Expert Systems with Applications*, 91: 127–137.

Gülçehre, Ç.; Denil, M.; Malinowski, M.; Razavi, A.; Pascanu, R.; Hermann, K. M.; Battaglia, P. W.; Bapst, V.; Raposo, D.; Santoro, A.; and de Freitas, N. 2018. Hyperbolic Attention Networks. *CoRR*, abs/1805.09786.

He, R.; Lee, W. S.; Ng, H. T.; and Dahlmeier, D. 2017. An unsupervised neural attention model for aspect extraction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 388–397.

He, X.; and Chua, T.-S. 2017. Neural factorization machines for sparse predictive analytics. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 355–364.

Hou, S.; Fan, Y.; Ju, M.; Ye, Y.; Wan, W.; Wang, K.; Mei, Y.; Xiong, Q.; and Shao, F. 2021. Disentangled Representation Learning in Heterogeneous Information Network for Large-scale Android Malware Detection in the COVID-19 Era and Beyond. *35th AAAI Conference on Artificial Intelligence (AAAI)*.

Hu, L.; Xu, S.; Li, C.; Yang, C.; Shi, C.; Duan, N.; Xie, X.; and Zhou, M. 2020. Graph Neural News Recommendation with Unsupervised Preference Disentanglement. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 4255–4264. Online: Association for Computational Linguistics.

Hu, M.; and Liu, B. 2004. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168–177.

Hu, Z.; Yang, Z.; Liang, X.; Salakhutdinov, R.; and Xing, E. P. 2017. Controllable Text Generation. *CoRR*, abs/1703.00955.

Huang, J.; Wang, Z.; Chang, K. C.; Hwu, W.; and Xiong, J. 2020. Exploring Semantic Capacity of Terms. *CoRR*, abs/2010.01898.

Iyyer, M.; Guha, A.; Chaturvedi, S.; Boyd-Graber, J.; and Daumé III, H. 2016. Feuding families and former friends: Unsupervised learning for dynamic fictional relationships. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1534–1544.

Jang, E.; Gu, S.; and Poole, B. 2017. Categorical Reparameterization with Gumbel-Softmax.

Jin, W.; Ho, H. H.; and Srihari, R. K. 2009. OpinionMiner: A novel machine learning system for web opinion mining and extraction. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1195–1204.

Karamanolakis, G.; Hsu, D.; Hsu, D.; and Gravano, L. 2019. Leveraging Just a Few Keywords for Fine-Grained Aspect Detection Through Weakly Supervised Co-Training. *CoRR*, abs/1909.00415.

Krioukov, D.; Papadopoulos, F.; Kitsak, M.; Vahdat, A.; and Boguñá, M. 2010. Hyperbolic geometry of complex networks. *Physical Review E*, 82(3).

Le, M.; Roller, S.; Papaxanthos, L.; Kiela, D.; and Nickel, M. 2019. Inferring Concept Hierarchies from Text Corpora via Hyperbolic Embeddings. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3231–3241.

Liu, B. 2012. Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1): 1–167.

Liu, P.; Joty, S.; and Meng, H. 2015. Fine-grained opinion mining with recurrent neural networks and word embeddings. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 1433–1443.

Liu, Q.; Nickel, M.; and Kiela, D. 2019. Hyperbolic Graph Neural Networks. *CoRR*, abs/1910.12892.

Liu, Y.; Wang, X.; Wu, S.; and Xiao, Z. 2020. Independence Promoted Graph Disentangled Networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04): 4916–4923.

López, F.; Heinzerling, B.; and Strube, M. 2019. Fine-Grained Entity Typing in Hyperbolic Space. *CoRR*, abs/1906.02505.

López, F.; and Strube, M. 2020. A Fully Hyperbolic Neural Model for Hierarchical Multi-Class Classification. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, 460–475. Online: Association for Computational Linguistics.

Lund, J.; Cook, C.; Seppi, K.; and Boyd-Graber, J. 2017. Tandem anchoring: A multiword anchor approach for interactive topic modeling. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 896–905.

Luo, L.; Ao, X.; Song, Y.; Li, J.; Yang, X.; He, Q.; and Yu, D. 2019. Unsupervised Neural Aspect Extraction with Sememes. In *IJCAI*, 5123–5129.

Ma, J.; Cui, P.; Kuang, K.; Wang, X.; and Zhu, W. 2019a. Disentangled graph convolutional networks. In *International Conference on Machine Learning*, 4212–4221. PMLR.

Ma, J.; Zhou, C.; Cui, P.; Yang, H.; and Zhu, W. 2019b. Learning Disentangled Representations for Recommendation. *CoRR*, abs/1910.14238.

Mitchell, M.; Aguilar, J.; Wilson, T.; and Van Durme, B. 2013. Open domain targeted sentiment. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 1643–1654.

Murty, S.; Verga, P.; Vilnis, L.; Radovanovic, I.; and McCallum, A. 2018. Hierarchical losses and new resources for fine-grained entity typing and linking. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 97–109.

Ni, J.; Li, J.; and McAuley, J. 2019. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP-IJCNLP)*, 188–197.

Nickel, M.; and Kiela, D. 2017. Poincaré embeddings for learning hierarchical representations. *arXiv preprint arXiv:1705.08039*.

Nickel, M.; and Kiela, D. 2018. Learning Continuous Hierarchies in the Lorentz Model of Hyperbolic Geometry. *CoRR*, abs/1806.03417.

Pontiki, M.; Galanis, D.; Papageorgiou, H.; Androutsopoulos, I.; Manandhar, S.; Al-Smadi, M.; Al-Ayyoub, M.; Zhao, Y.; Qin, B.; De Clercq, O.; et al. 2016. SemEval-2016 Task 5: Aspect based sentiment analysis. In *International Workshop on Semantic Evaluation*, 19–30.

Robbin, J. W.; and Salamon, D. A. 2011. Introduction to differential geometry. *ETH, Lecture Notes, preliminary version*, 18.

Sa, C. D.; Gu, A.; Ré, C.; and Sala, F. 2018. Representation Tradeoffs for Hyperbolic Embeddings. *CoRR*, abs/1804.03329.

Sala, F.; De Sa, C.; Gu, A.; and Ré, C. 2018. Representation tradeoffs for hyperbolic embeddings. In *International Conference on Machine Learning*, 4460–4469. PMLR.

Sarkar, R. 2011. Low distortion Delaunay embedding of trees in hyperbolic plane. In *International Symposium on Graph Drawing*, 355–366. Springer.

Shen, T.; Lei, T.; Barzilay, R.; and Jaakkola, T. 2017. Style transfer from non-parallel text by cross-alignment. *arXiv preprint arXiv:1705.09655*.

Shi, T.; Kang, K.; Choo, J.; and Reddy, C. K. 2018. Short-text topic modeling via non-negative matrix factorization enriched with local word-context correlations. In *Proceedings of the 2018 World Wide Web Conference*, 1105–1114.

Shi, T.; Li, L.; Wang, P.; and Reddy, C. K. 2020. A Simple and Effective Self-Supervised Contrastive Learning Framework for Aspect Detection. arXiv:2009.09107.

Srivastava, A.; and Sutton, C. 2017. Autoencoding variational inference for topic models. *arXiv preprint arXiv:1703.01488*.

Sun, J.; Cheng, Z.; Zuberi, S.; Pérez, F.; and Volkovs, M. 2021. HGCF: Hyperbolic Graph Convolution Networks for Collaborative Filtering.

Tai, C.-Y.; Huang, C.-K.; Huang, L.-Y.; and Ku, L.-W. 2021. Knowledge Based Hyperbolic Propagation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1945–1949.

Tay, Y.; Luu, A. T.; and Hui, S. C. 2017. Enabling Efficient Question Answer Retrieval via Hyperbolic Neural Networks. *CoRR*, abs/1707.07847.

Tulkens, S.; and van Cranenburgh, A. 2020. Embarrassingly Simple Unsupervised Aspect Extraction. *CoRR*, abs/2004.13580.

Wang, W.; Pan, S. J.; Dahlmeier, D.; and Xiao, X. 2016. Recursive neural conditional random fields for aspect-based sentiment analysis. *arXiv preprint arXiv:1603.06679*.

Wang, X.; Jin, H.; Zhang, A.; He, X.; Xu, T.; and Chua, T.-S. 2020. Disentangled Graph Collaborative Filtering. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1001–1010.

Xu, P.; and Barbosa, D. 2018. Neural fine-grained entity type classification with hierarchy-aware loss. *arXiv preprint arXiv:1803.03378*.

Yang, B.; and Cardie, C. 2012. Extracting opinion expressions with semi-Markov conditional random fields. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing*, 1335–1345.

Yoshua Bengio, A. C. C.; and Vincent, P. 2012. Unsupervised Feature Learning and Deep Learning: A Review and New Perspectives. *CoRR*, abs/1206.5538.

Zhang, C.; and Gao, J. 2020. Hype-HAN: Hyperbolic Hierarchical Attention Network for Semantic Embedding. In *IJCAI*, 3990–3996.

Zhao, J.; Kim, Y.; Zhang, K.; Rush, A.; and LeCun, Y. 2018. Adversarially regularized autoencoders. In *International Conference on Machine Learning*, 5902–5911. PMLR.

Zhuang, H.; Guo, F.; Zhang, C.; Liu, L.; and Han, J. 2020. Joint Aspect-Sentiment Analysis with Minimal User Guidance. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1241–1250.