

Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review

Rostoucí objem obsahu vytvářeného uživateli na webu učinil z analýzy sentimentu důležitý nástroj pro extrakci informací o lidském emocionálním stavu.

Ve své stěžejní práci na Aspect-based Sentiment Analysis (ABSA), Hu a kol. (2004) tvrdili, že studium analýzy sentimentu je možné na třech úrovních – dokument, věta a entita nebo aspekt. Zaměření na úroveň dokumentu nebo věty předpokládá, že v dokumentu nebo větě je vyjádřeno pouze jedno téma, což v mnoha situacích neplatí. Důkladnější analýza proto vyžaduje prošetření na úrovni subjektů a aspektů, aby bylo možné identifikovat subjekty a související aspekty a klasifikovat pocity spojené s těmito subjekty a aspekty. Pro účely tohoto dokumentu ABSA znamená analýzu sentimentu na úrovni entity nebo aspektu.

Tento druh jemnozrné analýzy se obecně opíral o techniky strojového učení, které, i když jsou účinné, vyžadují velké, doménově specifické datové sady a manuální trénovací data (Hu & Liu, 2004). Kromě toho může být aspekt reprezentován různými slovy vyžadujícími více než jeden klasifikační algoritmus (Schouten & Frasincar, 2016). V poslední době se slibně ukázala experimentální práce s metodami strojového učení, kdy Poria, Cambria a Gelbukh (2016) uvádějí vyšší přesnost pomocí hlubokých konvolučních neuronových sítí, což je vlastnost hlubokého učení (DL), pojmenovaná pro svou „hlubokou“ techniku vícevrstvého zpracování, která používá po sobě jdoucí modulové vrstvy k vybudování na předchozím výstupu pomocí algoritmu zpětného šíření (Lecun, Bengio & Hinton, 2015). V každé vrstvě je vstup převeden na číselná vyjádření, která jsou následně klasifikována. Dosahuje se tak stále vyšší úrovně abstrakce (Goodfellow, Bengio & Courville, 2016). Řada algoritmů (tj. hluboké neuronové sítě (DNN), rekurentní neuronové sítě (RNN), konvoluční neuronové sítě (CNN), rekursivní neuronové sítě (RecNN) atd.) usnadňují analýzu v různých oblastech pomocí hlubokých neuronových sítí zvláště vhodných pro jemné - zrnitá práce díky značnému počtu vrstev připojených procesorů, aktivovaných buď senzory z prostředí nebo váženými výpočty z předchozích neuronů (Schmidhuber, 2015).

Rané přístupy k DL zkoumaly lingvistické rysy, gramatické vztahy, klasifikátory strojového učení a tématické modelování k identifikaci aspektů a polarit (Schouten & Frasincar, 2016). V poslední době byly metody DL úspěšně aplikovány na zpracování přirozeného jazyka (NLP), díky čemuž je zajímavé zkoumat, jak si DL vedl, když byly nastaveny jemnozrné úkoly, jako je ABSA

Tento dokument, spíše než opakování zjištěných zjištění z předchozích průzkumů, si klade za cíl představit a porovnat novější vývoj v přístupech DL obecně a pro ABSA zvláště. Tato recenze je speciálně navržena pro studenty a výzkumníky v oblasti zpracování přirozeného jazyka, kteří by rádi prozkoumali hluboké neuronové sítě a také nejnovější trendy ve výzkumu v ABSA.

Aspect-Based Sentiment Analysis

Pontiki a kol. (2016) přidělili ABSA tři důležité dílčí úkoly: (i) Extrakce cíle (OTE), (ii) Detekce kategorie aspektů (ACD) a (iii) Sentiment Polarity (SP), přičemž OTE se zabývá extrakcí

aspektů. (tj. entita nebo atribut), ACD s identifikací přidružených entit a atributů a (iii) a SP s objasněním polarity sentimentu aspektů.

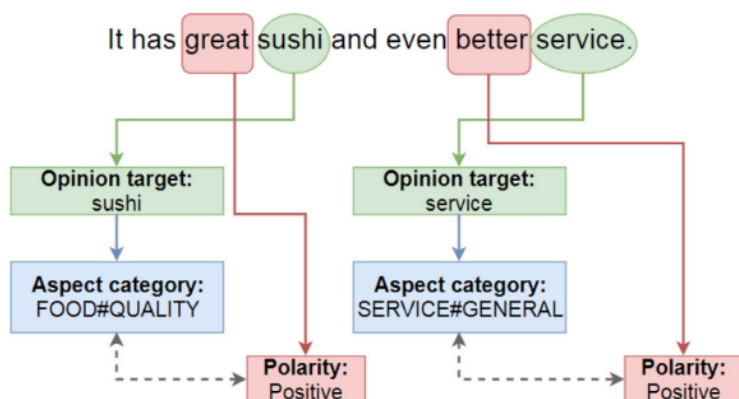


Fig. 1. Three tasks of ABSA in a sample sentence from SemEval ABSA dataset 2016. The sentence has two opinion targets: sushi & service. The category of "sushi" is "Food", with the attribute being "Quality" and polarity "Positive". The category is "Service", with an attribute of "General" and polarity of "Positive".

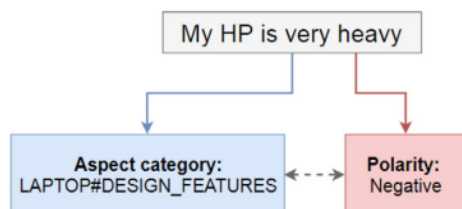


Fig. 2. The opinion target is not explicitly stated, but the category and polarity of the sentence can still be inferred. Sentence from SemEval data 2016.

ABSA se používá hlavně na zákaznické recenze z webových stránek a platformem elektronického obchodu, jako jsou Amazon, Yelp, Taobao a další. Pravděpodobně se jedná o recenze produktů nebo služeb a lze předpokládat, že v každém z nich je zmíněn pouze jeden subjekt, ale jeden nebo více aspektů (Saeidi, Bouchard, Liakata & Riedel, 2016). V posledních letech byly vyvinuty systémy pro oblasti, jako jsou elektronické recenze produktů (notebook, fotoaparát a telefon) a recenze pohostinství (restaurace, hotely). Byla zpřístupněna řada referenčních datových sad, včetně datové sady zákaznických recenzí od Hu et al. (2004) a řadu datových sad vydaných „International Workshop on Semantic Evaluation“ (SemEval 2014–2016) o recenzích notebooků, fotoaparátů, restaurací a hotelů (Pontiki et al., 2014, 2015, 2016). Další linií výzkumu pro ABSA je cílená (neboli na cíli závislá) analýza sentimentu (Vo & Zhang, 2015), která klasifikuje názorové polarity určité cílové entity uvedené ve zkoumaných větách (obvykle tweet).

Dřívější přístupy k identifikaci OTE a ACD (např. Hu et al., 2004) byly založeny na četnosti podstatných jmen a jmenných frází v textu s předpokladem, že aspektová slova se budou častěji opakovat. Omezením tohoto přístupu je závislost na frekvenci určitých kategorií slov (podstatná jména/podstatné fráze), což může dobře fungovat, pokud text obsahuje termíny s vysokou frekvencí, ale může selhat, pokud termíny nejsou časté (Rana et al., 2016).

Jiní extrahovali OTE a ACD využitím názorových a cílových vztahů. Poria, Chaturvedi, Cambria a Bisio (2016) a Piryani, Gupta a Singh (2017) se zaměřili na lingvistické vzorce založené na pravidlech, včetně zastavovacích slov a negace atd. Zde se předpokládalo, že

je snazší odhalit sentiment než aspekt slova. Autoři navrhli soubor názorových pravidel, aby nejprve identifikovali sentimentální slovo a poté použili gramatické vztahy k sestavení syntaktické struktury vět a ke zjištění aspektu. Poslední krok sestává z upřesnění, kde jsou přidána málo častá slova a odstraněny irelevantní aspekty. Klíčovým prvkem této metody, která je schopna identifikovat nízkofrekvenční aspekty, je lexikální vztah mezi sentimentovými slovy a aspekty (Schouten et al., 2016). Nevýhodou je však spoléhání se na gramatickou správnost věty a požadavek na manipulaci (Poria, Cambria et al., 2016).

Mezinárodní workshopy o sémantickém hodnocení podporují rozvoj analýzy sentimentu na úrovni aspektů (Pontiki et al., 2014, 2016, 2015), která poskytuje metodologii řízeného hodnocení a sdílené datové soubory pro všechny účastníky. Pro měření účinnosti klasifikačního modelu byla navržena čtyři hlavní měření: Precision (P), Recall (R), F-score (F1) a Accuracy (Acc).

Hluboké učení

(DL) je metoda strojového učení založená na reprezentaci dat učení pomocí algoritmů – umělých neuronových sítí a sítí přesvědčení – založených na více vrstvách modulů, kde je vstup analyzován a klasifikován, přičemž výstup z jedné vrstvy je přiváděn do další vrstvy jako vstup. Tento proces je známý jako backpropagation (Lecun et al., 2015), přičemž aktivace iniciuje zpětný výpočet gradientu objektivní funkce (Schmidhuber, 2015). Bez ohledu na typ vstupu (tj. zvuk, obrázek nebo text), je převeden na číselné vektory a poté seskupen do smysluplné klasifikace. Protože každá následující vrstva odpovídá zvýšené úrovni abstrakce, lze říci, že DL představuje „vnořené hierarchie“ jednodušších konceptů (Goodfellow et al., 2016). Dalším rysem je, že jeho úroveň hloubky lze považovat za podobnou vícefázovému programování, ve kterém je každá vrstva stavem paměti počítače po provedení sady instrukcí (Goodfellow et al., 2016). Zvýšením úrovně hloubky se zvyší schopnost systému selektivně a invariantně reprezentovat (Lecun, Bengio, & Hinton, 2015).

Hluboké neuronové sítě (DNN) jsou dobrými příklady DL a jsou předmětem tohoto článku. Hluboké neuronové sítě jsou typy umělých neuronových sítí (algoritmů), které zahrnují značný počet vrstev „neuronů“ nebo připojených procesorů, aktivovaných buď senzory z okolí, nebo váženými výpočty z předchozích neuronů (Schmidhuber, 2015). Pro DL, stejně jako pro přístupy ke strojovému učení obecně, jsou datové sady často rozděleny do tří složek: tréninkové, validační a testovací datové sady, které jsou v souladu s obecnými principy strojového učení. Lecun a kol. (2015) zmapovali tréninkový proces jako převod vstupu do vektorového skóre bez ohledu na typ vstupu (tj. obrázky). Zpočátku se objeví chybové skóre, které je třeba snížit trénováním algoritmu tak, aby přesněji odpovídal nastaveným parametrům (váhám) pro cílové slovo (nebo obrázek). Úpravy jsou nahrazeny strojem, aby se snížila chyba. Spouštěč pro úpravu chyby je „gradientový vektor“, který reaguje na manipulaci s parametry, které potřebují vyvážit chybu.

Přístupy hluboké neuronové sítě (DNN) k NLP se vyznačují (i) hustým vkládáním slov; (ii) více skrytých vrstev mezi vstupem a výstupem; a (iii) výstupní jednotky (obr. 3). Vložení slov jsou reprezentace slov v d-rozměrném prostoru, zakódované jako husté numerické vektory (Rojas-Barahona, 2016). Tyto vektory, tvrdí Levy a Goldberg (2014), určují pravděpodobnost, že se slovo objeví ve specifické matici slov (tj. s přidruženými slovy). Jedním z prvních modelů vkládání slov byl model Bengia et al. (2006), který navrhl neurální

pravděpodobnostní jazykový model se sdílenou vyhledávací tabulkou. V nedávných modelech DNN jsou vkládání slov obvykle předem natrénovaná, ale ne specifická pro úkol, takže vektory učících se slov mohou zachytit obecné syntaktické a sémantické informace (Chen, Xu, He, & Wang, 2017; P. Liu et al., 2015; Poria, Cambria, et al., 2016). Existují různé modely pro vkládání slov, jako je Word2Vec (Mikolov, Corrado, Chen, & Dean, 2013), které kódují kontextové informace pomocí kontinuálních Bag-of-Words (CBOW) a modelů přeskakujících gramů. Vkládání slov je dále popsáno v části 3.3.1. Druhý rys – skryté vrstvy – lze konstruovat v různých formách a architekturách, tedy dopředné sítě a rekurzivní nebo rekurzivní sítě (Goldberg, 2016). Každá skrytá vrstva je složena z více neuronů, naskládaných dohromady pro výpočet nelineárních výstupů (Lecun et al., 2015). Obecně platí, že vyšší vrstvy se vyvíjejí prostřednictvím školení, aby využívaly komplexní kompoziční nelineární funkce nižších vrstev, a tudíž zachycovaly abstraktnější reprezentace než nižší vrstvy (Goldberg, 2016).

V níže uvedených částech budou přezkoumány hlavní modely DNN aplikované na úlohy ABSA, včetně konvolučních neuronových sítí (CNN), rekurentních neuronových sítí (RNN), rekurzivních neuronových sítí (RecNN) a hybridních modelů v oddílech 3.2 - 3.5. V každé sekci bude uveden přehled architektury, po kterém následuje aplikace modelů na úlohy ABSA.

Convolutional Neural Network Model

CNN se stala oblíbeným modelem DL mezi výzkumníky NLP, protože průkopnické práce Collobert et al. (2011) a Kim (2014), kteří obhajovali úspěch CNN v řadě úkolů NLP, včetně analýzy sentimentu. Hlavní předností CNN je její schopnost extrahovat ze vstupu nejdůležitější n-gramové rysy a vytvořit „informativní latentní sémantické znázornění“ pro provádění dalších klasifikačních úkolů (Rojas-Barahona, 2016; Young et al., 2017; L Zhang a kol., 2018). Základní jednovrstvý CNN pro modelování vět může sestávat ze 4 vrstev, jak je uvedeno

Vstupní vrstva : představující větu délky n jako $x = x_1 \& x_2 \& \dots \& x_n$ kde $x_i \in \mathbb{R}$ je d -rozměrný vektor slova odpovídající i -tému slovu ve větě a je zřetězením operátor. Konvoluční vrstva : tato vrstva generuje novou vlastnost c s filtrem $w \in \mathbb{R}^{h \times k}$ pomocí okna h slov od i do $i + h - 1$. Motivací pro využití modelu CNN v úlohách ABSA je předpoklad, že klíčová slova mohou obsahovat výraz aspektu a označují kategorii nebo určují polaritu bez ohledu na jejich polohu. CNN se svou architekturou (obr. 7) dokáže naučit tyto funkce nacházet, a je tedy schopna extrahovat místní vzory z dat bez ohledu na jejich umístění. To je velmi užitečné pro identifikaci frází s pevnou délkou (Goldberg, 2017). Další výhodou je, že CNN je nelineární model, u kterého se očekává, že bude lépe odpovídat datům než lineární modely, jako je CRF, a nevyžaduje rozsáhlé ručně vytvořené prvky, jako jsou pevná jazyková pravidla (Poria, Cambria a kol., 2016). Model CNN byl aplikován na všechny úkoly ABSA, zejména v doméně spotřebitelského hodnocení (tabulka 4). Pro úkol OTE je hlavním příkladem úspěšných studií Poria, Cambria a kol. (2016), kteří přizpůsobili architekturu CNN z větné reprezentace na predikci založenou na slovech. Za předpokladu, že značka každého slova je závislá na každém sousedním slově, vytvořili kolem každého slova ve větě okno lokálních vlastností pěti slov.

Recurrent neural network

The simple RNN model is based on the Elman network (Elman,

1991; Goodfellow et al., 2016) with direct cycles in their hidden connection (Goldberg, 2016; Rojas-Barahona, 2016). This model proposes that the hidden state is dependent on the input and past hidden state, with the same function and the same set of parameters being used at every time step.

Rekurentní neuronové sítě se staly populárními v úlohách analýzy sentimentu. Základem modelů RNN je, že vektor s pevnou velikostí představuje jednu sekvenci (tj. větu nebo dokument) tím, že každý token vloží do opakující se jednotky, takže může zachytit inherentní sekvenční povahu jazyka (tj. jedno slovo rozvíjí jeho sémantický význam díky svému předchozímu slovu) (Goldberg, 2016; Goodfellow et al., 2016). Ve srovnání s modely CNN mají modely RNN flexibilní výpočetní kroky, že výstup z RNN je závislý na předchozích výpočtech, takže je schopen zachytit kontextové závislosti v jazyce a také modelovat různé délky textu (Tang, Qin a Liu, 2016). Model RNN má ve srovnání s dopřednou neuronovou sítí dvě důležité vlastnosti. Za prvé, na rozdíl od CNN má různé parametry v každé vrstvě, parametry v RNN jsou stejné v každém kroku, což pak snižuje počet parametrů potřebných k učení (L. Zhang et al., 2018). Za druhé, protože výstup jednoho stavu závisí na předchozím stavu, lze říci, že RNN má paměť předchozích výpočtů, díky čemuž je lepší ve zpracování sekvenčních informací ve srovnání s CNN. Tři výše uvedené modely se zaměřují na použití minulých slov k předpovědi dalšího slova. V praxi by mnoho studií chtělo provádět předpovědi na základě budoucích slov, a proto jsou navrženy obousměrné modely RNN se začleněním dopředné a zpětné vrstvy, aby bylo možné získat informace z předchozích a následujících tokenů (Graves, 2008; Fan, Qian, Xie, & Soong, 2014). Ve srovnání s jinými modely neuronových sítí mají RNN a obousměrné RNN obrovské zastoupení v klasifikaci na úrovni slov a na úrovni vět v doméně spotřebitelských recenzí (Tabulka 6). Mnoho modelů založených na RNN využilo obousměrného RNN k překódování minulých a budoucích kontextů. Přístup od Chaudhuri & Ghosh (2016) použil hierarchické obousměrné RNN v ACD ve vysoce zkrácených datech domény hodnocení hotelů a získal lepší výsledek než LSTM. Model se skládá ze šesti vrstev včetně čtyř vrstev obousměrných RNN, jedné plně propojené vrstvy a jedné softmaxové vrstvy.

Jedním z nejúspěšnějších pokusů je zkombinovat RNN s CRF klasifikační vrstvou, takže model nejen zachycuje dlouhodobou závislost celých vět, ale také využívá závislost každého štítku na sobě

Modely RNN jsou také aplikovány v doméně Twitteru a nových com- (Tabulka 7) spíše než model CNN, jak je vysvětleno výše na omezení modelu CNN v zachycení dlouhodobých závislostí. Un- stejně jako doména zákaznických recenzí je napadena doména Twitteru s omezenou délkou, neformálními kontexty a používáním emocí záporny (Giachanou & Crestani, 2016). Nicméně studie využívající RNN modely prokázaly v tomto úkolu konkurenční výkon

Modely rekurzivní neuronové sítě (RecNN)

jsou lingvisticky motivovány tím, že zkoumají stromové struktury (např. syntaktické struktury) a jejich cílem je naučit se elegantně kompoziční sémantiku. Přirozený jazyk pravděpodobně vykazuje přirozenou rekurzivní strukturu, umisťuje slova a fráze hierarchickým způsobem. Stromově strukturované modely tak mohou lépe využívat takové syntaktické interpretace větné struktury (Socher, Perelygin, & Wu, 2013; Tai, Socher, & Manning, 2015). Obecně platí, že v rekurzivní neuronové síti je vektorová reprezentace každého uzlu ve stromové struktuře vypočítána z reprezentace všech jeho potomků pomocí váhové matice W , která je sdílena v celé síti (Socher et al., 2013).

Hybridy

Mnoho studií se vyrovnalo s výhodami i nevýhodami dříve diskutovaných modelů a pokusilo se aplikovat hybridní řešení v doménách zákaznických recenzí, jako jsou Xue, Zhou, Li, & Wang (2017), Ye et al. (2017), Chen, Xu, Yang a Li (2016). Xue a kol. (2017) poznamenali, že termíny aspektů a kategorie aspektů spolu úzce souvisejí, a proto navrhli víceúlohový rámec BiLSTM pro OTE a CNN pro ACD. Hlavními výhodami tohoto rámce je vzájemné sdílení informací o dvou úlohách, ve kterých může CNN také využít další informace získané v BiLSTM ke zlepšení svých informačních funkcí, zatímco predikovaný tag z BiLSTM může také přijímat nejvýraznější n -gramy. funkce prostřednictvím konvolučních operací. Podobně Chen a kol. (2016) také zkombinovali LSTM a CNN dohromady pro klasifikaci pocitů, ale použili LST pro generování vkládání kontextu a CNN pro detekci prvků. Ye a kol. (2017) navrhli konvoluční vrstvenou neuronovou síť (DTBCSNN) založenou na stromech závislostí pro extrakci termínů aspektů, ve které je konvoluce zahrnuta do stromů analýzy závislostí věty, aby zachytila syntaktické a sémantické rysy. To může překonat praktická omezení sekvenčních modelů (RNN), které nedokážou zachytit stromové informace o závislostech. Navrhovaný model nepotřebuje žádné rukodělné rysy a je flexibilní, aby zahrnoval další jazykové vzory. Tabulka 9 níže poskytuje přehled hybridních řešení.

Comparison

Toto pozorování je potvrzeno srovnáním modelů. Je jasné, že modely CNN mohou být vysoce efektivní (Kim, 2014; Poria, Cambria, et al., 2016; Wu et al., 2016) díky své schopnosti extrahovat lokální vzory (tj. nejdůležitější n -gram) věty pro vytvoření vstupu s pevnou velikostí. To však platí pouze tehdy, pokud je vyžadována klasifikace klíčových frází omezené délky (Goldberg, 2017). Modely CNN navíc vyžadují velké soubory tréninkových dat a vyžadují značné množství přesně vyladěných parametrů.

RNN jsou výkonné, protože kombinují dvě vlastnosti: (i) Distribuované skryté stavy, které jim umožňují efektivně ukládat informace z minulých výpočtů; a (ii) Nelineární dynamika, která lépe odpovídá nelineární povaze dat (Tang, Qin, Feng, et al., 2015; Y. Wang et al., 2016).

V případě RecNN jsou jednoznačnými výhodami jednoduchá architektura a schopnost učit se stromové struktury vět a nová slova (Socher et al., 2013; Tai et al., 2015). Jsou však silně závislé na analyzátorech (Socher et al., 2013) a dosud nevykazovaly konzistentní výkon při klasifikaci vět (Goldberg, 2017). Je zřejmé, že je zapotřebí další výzkum.

Tyto různé modely byly tedy navrženy s různými cíli, pokud jde o modelování vět, zejména při analýze CNN a RNN. Zatímco CNN se snaží extrahovat nejdůležitější n-gramy, RNN se snaží vytvořit kompozici s neomezeným kontextem (Young et al., 2017; L. Zhang et al., 2018)