Concept-to-code: Aspect Sentiment Classification with Deep Learning

Muthusamy Chelliah
Flipkart
muthusamy.c@flipkart.com

Asif Ekbal
IIT Patna
asif@iitp.ac.in

Mohit Gupta
Flipkart
mohit.gupta@flipkart.com

ABSTRACT
Aspect sentiment classification (ASC) is more fine-grained than document- or sentence- level tasks in sentiment analysis. Neural networks alleviate feature engineering and attention mechanism in particular addresses targeted-context detection problem. LSTM and memory networks are 2 models which incorporate attention in recent literature for ASC. This tutorial is an advanced survey equally of interest to academic researchers and industry practitioners - very timely with so much vibrant research in the NLP community over the past 5 years. We not only review relevant concepts from papers across multiple research groups (with 1 out of 20 papers listed co-authored by us) but also present code fragments which illustrate such techniques and could be leveraged in due course in use cases from online marketplaces like Flipkart where product reviews influence user purchases.

TYPE OF TUTORIAL:
- Survey, introducing expert, non-specialist to sentiment analysis

TUTORIAL OUTLINE (180 min):
1. INTRODUCTION: (15min)
- E-commerce customers are interested not just in overall product ratings but also detailed feature/ function/attributes of products. Fine-grained opinion analysis helps detect subjective expressions in a text (e.g. hate), characterize their intensity (e.g. strong) and sentiment (e.g. negative) as well as to identify target/ topic (i.e. what the opinion is about). Aspect-based sentiment analysis (ABSA) itself thus has 3 sub-tasks in entity (e.g., restaurant) review mining: term extraction (e.g., space), category detection(e.g., ambience) and polarity detection (e.g., narrow (negative)). Given an aspect (aka opinion target) and a sentence, aspect-based sentiment classification (ASC) of polarity (e.g., positive, neutral, negative) expressed on target in the sentence thus is focus of this tutorial.

- Mining features automatically from online reviews is a key step for opinion summarization. Opinion words co-occur with aspect words and hence provide indicative clues for aspect detection. Given a predefined set, classification algorithm can leverage hand-crafted features to identify aspect category in a review sentence. Human effort required by feature engineering makes this approach unstable when product domain changes. A sentence could contain multiple sentiment-target pairs; thus the main challenge of ASC is to separate different opinion contexts for different targets.

- Now that we have introduced the problem more from business/ customer perspective, we will give a quick, technical overview of deep learning applications in ABSA next.

2. BACKGROUND: (30min)
- Distributed representation learners (e.g., recurrent neural network (RNN)) alternatively model latent features as dense vectors of hidden layers; RNNs can operate on sequential data (e.g., label) of variable length too. Bidirectional RNNs incorporate information from preceding as well as following tokens while word embedding induction has enabled effective training of RNNs by allowing a lower dimensional dense input representation and hence, more compact networks.
- A general class of discriminative models based on RNNs and word embeddings help opinion target identification without task-specific feature engineering [Liu 15]. Pre-trained word embeddings from external sources help initialize RNNs which then fine-tune word vectors during training to learn task-specific embeddings. Lexical semantic clue verifies whether a candidate term is related to target product through a similarity graph; a convolutional neural network model (CNN) then captures contextual semantic clue; label propagation combines both clues [Xu 14]. A semi-supervised word embedding algorithm first obtains continuous word representations on a large set of reviews with noisy labels towards aspect category detection [Zhou 15]; a logistic regression classifier is trained next on deeper/hybrid features generated through neural networks stacked on word vectors.
- Once we set context about the broader ABSA problem thus so far, we next dive deep into one of the two sub-tasks: ASC; comprehensive review of extraction of aspect/sentiment (terms) briefly outlines here is thus outside the scope of this proposal.

3 ASPECT SENTIMENT CLASSIFICATION (30 min)
3.1 RECURRENT NEURAL NETWORK:
- 3.1 RNNs help identify phrase/sentence sentiment hierarchically using syntactic parse trees [Socher 13]. Binary dependency tree represents syntactic relations associated with target aspect in a sentence with each word represented as multi-dimensional vectors; target-dependent phrase tree is constructed by combining constituent/ dependency trees using two kinds of composition, instead of a list of global functions [Nguyen 15]. Finally, deep RNN with multiple stacked hidden layers, naturally employ a temporal hierarchy with multiple layers operating at different time scales: lower levels capture short term interactions among words; higher layers reflect interpretations aggregated over longer spans of text [Irsoy 14].

3.2 LSTM AND ATTENTION (40 min)
- LSTM is sequential, manipulates each word with same operation and captures context information in an implicit way. Attention mechanism helps capture importance of each context word towards a target by modeling their semantic associations. Simply mapping a target by averaging its component word vectors may work fine for targets that only contain one word but may fail to capture semantics of more complex expressions (e.g., hot dogs).

- Two ways are proposed to incorporate aspect information during attention in an LSTM [Wang 16]: concatenate aspect vectors into sentence hidden representations for computing attention weights and append aspect vector into input word vectors.

- Attending only to sentiment information (e.g., great) and ignoring aspect-related information (e.g., tastes) may cause mismatch when an unrelated sentiment word (e.g., dreadful) is semantically meaningful for the given aspect (e.g., food). A mask layer is introduced in [Cheng 17] to represent location information of aspect terms and sentiment expressions, which improves attention calculation of sentiment expressions in the hierarchical attention network which supports to extract aspect terms together with aspect-level sentiment classification.

- A method for target representation that better captures semantic meaning of opinion target is proposed first [He 18]. Next, an attention model that incorporates syntactic information into the attention mechanism is proposed. For aspect extraction, probability distribution over aspects for given target is learnt, and weighted summation of aspect embeddings for target representation is used.

- Adjacent context words (e.g., wonderful) may be far away in reality from target word (e.g., battery life). [Li 18] introduces position embeddings to learn position-aware representations of sentences and further generate target-specific representations of contextual words. Hierarchical attention-based mechanism fuses information of targets and contextual words finally.

- There is a need for incorporating importance degrees of both words and clauses inside a sentence. Multiple bi-directional LSTM layers are leveraged to encode all clauses in the sentence [Wang 18] and a word-level attention layer captures importance degrees of words in each clause. Another Bidirectional LSTM layer encodes output from former layers and a clause-level attention layer capture importance degrees of all clauses inside a sentence.

- Global attention suffers from assigning high score to irrelevant sentiment words where sentence contains noisy words or multiple targets. A global attention is employed to capture coarse information about the target, and a syntax-directed local attention instead [Wang 19] is used to attend to words syntactically close to the target. An information gate is then utilized to synthesize the information from local and global attention results for a more sentiment-oriented representation.

3.3 MEMORY NETWORKS (40 min)

- A multi-layered approach with shared parameters - where each layer is a content- and location-based attention model, learns importance/ weight of each context word and calculates continuous text representation [Tang 16].

- A deep memory network interleaves both the tasks in ABSA [Li 17]. Signals produced in target detection provide clues for polarity classification, and reversely, the predicted polarity provides feedback to the identification of targets. Treatments for the set of targets also influence each other – learned representations may share same semantics for some targets but vary for others. Two LSTMS equipped with extended memories and neural memory operations are designed for jointly handling aspect and opinion extraction.

- Performance of memory networks simply relying on attention degrades when the sentiment of a context word (e.g., price) - unlike target-independent context words (e.g., ridiculous, excellent) is sensitive to the given target (e.g., screen resolution vs. price). Target-sensitive memory networks (TMNs) [Wang 18] can capture sentiment interaction between targets and contexts.

- A similarity model [Tay 18] learns to attend based on associative relationships (i.e., circular convolution) between sentence words and aspect thus alleviating flaws of naive concatenation (e.g., implications of doubled input/parameter costs to LSTM layer on memory footprint, computational complexity and risk of overfitting).

- As specific instances of aspects, terms explicitly occur in sentences. It is beneficial for models to focus on nearby context words. In contrast, as high level semantic concepts of terms (e.g., space), aspects (e.g., ambience) usually have more generalizable representations. However, conventional methods cannot utilize information of aspects and terms simultaneously, due to lack of annotations with both aspects (e.g., food) and terms (e.g., treats). A main memory is thus used to capture important context words (e.g., narrow, barely there, dirt-cheap) [Zhu 18]; an auxiliary memory implicitly convert aspects and terms to each other, and feeds both of them to the main memory.

- Sentiment (e.g., better deal) of an aspect (e.g., coffee) in a sentence can influence the succeeding aspects (e.g., sandwich) due to the presence of conjunctions (e.g., than). Aspects when arranged as a sequence thus reveal high correlation and interplay of sentiments. Recurrent memory networks leverage multihop attention in [Majumder 18] to refine target aspect representation incorporating other aspects.

4 FUTURE DIRECTIONS (20min)

- Capture word-level information lacks the capacity for modeling complicated expressions which consist of multiple words. A convolutional memory network [Fan 18] which incorporates an attention mechanism sequentially computes the weights of multiple memory units corresponding to multi-words.

- CNNs are good at capturing local patterns. Two neural units that take target aspects into account in a CNN-based approach [Huang 18] are parameterized filter and gate. Both are generated from aspect-specific features and are further applied on sentence.

- Handling sentiment negation, better embedding for multi-word phrases, analyzing sentiment composition, and learning better attention are future direction in aspect-sensitive sentiment classification.

5 **CONCLUSION (5min)**

- Despite utilising syntax structure of sentences, RNNs (e.g., LSTM) find it challenging to discriminate between different sentiment polarities at a fine-grained aspect level.
- Aspect embedding and attention weights help focus on different parts of a sentence.
- Faster/simpler deep memory networks capture of importance of context words.

REFERENCES:

[Socher 13] Socher R, Perelygin A, Wu J, Chuang J, Manning CD, Ng A, Potts C. Recursive deep models for semantic compositionality over a sentiment treebank.EMNLP 2013.

[Xu 14] Xu L, Liu K, Lai S, Zhao J. Product feature mining: Semantic clues versus syntactic constituents. ACL 2014.

[Irsoy 14] Irsoy O, Cardie C. Opinion mining with deep recurrent neural networks. EMNLP 2014.

[Nguyen 15] Nguyen TH, Shirai K. Phrasernn: Phrase recursive neural network for aspect-based sentiment analysis. EMNLP 2015.

[Liu 15] Liu P, Joty S, Meng H. Fine-grained opinion mining with recurrent neural networks and word embeddings. EMNLP 2015.

[Tang 16] Tang D, Qin B, Liu T. Aspect level sentiment classification with deep memory network. EMNLP 2016.

[Wang 16] Wang Y, Huang M, Zhao L. Attention-based lstm for aspect-level sentiment classification. EMNLP 2016.

[Cheng 17] Cheng J, Zhao S, Zhang J, King I, Zhang X, Wang H. Aspect-level sentiment classification with heat (hierarchical attention) network. CIKM 2017.

[Li 17] Li C, Guo X, Mei Q. Deep memory networks for attitude identification. WSDM 2017.

[Wang 18] Wang S, Mazumder S, Liu B, Zhou M, Chang Y. Target-sensitive memory networks for aspect sentiment classification, ACL 2018.

[Zhu 18] Zhu P, Qian T. Enhanced Aspect Level Sentiment Classification with Auxiliary Memory. COLING 2018.

[He 18] He R, Lee WS, Ng HT, Dahlmeier D. Effective attention modeling for aspect-level sentiment classification. COLING 2018.

[Tay 18] Tay Y, Tuan LA, Hui SC. Learning to attend via word-aspect associative fusion for aspect-based sentiment analysis. AAAI 2018.

[Li 18] Li L, Liu Y, Zhou A. Hierarchical Attention Based Position-aware Network for Aspect-level Sentiment Analysis. CoNLL 2018.

[Majumder 18] Majumder N, Poria S, Gelbukh A, Akhtar MS, Cambria E, Ekbal A. IARM: Inter-Aspect Relation Modeling with Memory Networks in Aspect-Based Sentiment Analysis. EMNLP 2018.

[Wang 18] Wang J, Li J, Li S, Kang Y, Zhang M, Si L, Zhou G. Aspect Sentiment Classification with both Word-level and Clause-level Attention Networks. IJCAI 2018

[Fan 18] Fan C, Gao Q, Du J, Gui L, Xu R, Wong KF. Convolution-based Memory Network for Aspect-based Sentiment Analysis. SIGIR 2018

[Yang 18] Yang J, Yang R, Wang C, Xie J. Multi-Entity Aspect-Based Sentiment Analysis with Context, Entity and Aspect Memory. AAAI 2018.

[Huang 18] Huang B, Carley K. Parameterized Convolutional Neural Networks for Aspect Level Sentiment Classification. EMNLP 2018

[Wang 19] Wang X, Xu G, Zhang J, Sun X, Wang L, Huang T. Syntax-Directed Hybrid Attention Network for Aspect-Level Sentiment Analysis. IEEE Acess, Jan. 2019

TARGET AUDIENCE

Researchers / industry practitioners with Computer Science / Statistics background; exposure to Machine Learning techniques desirable .

PREREQUISITE KNOWLEDGE OR SKILLS

Have exposure to Neural Network and Deep Neural network based techniques. Have applied understanding of machine learning terminologies and have gone through few data science projects end to end.

NAME, EMAIL AND BIO OF PRESENTERS:

Muthusamy Chelliah:
Email: muthusamy.c@flipkart.com
Address: Flipkart, Bangalore, India

Muthusamy Chelliah heads external research collaboration for Flipkart – who is a pioneer in the nascent online shopping market in a vibrant, emerging economy (India). He holds a PhD degree in Computer Science from Georgia Tech., Atlanta with a focus in distributed systems. He then spent 15 years with HP as a scientist and architect in US and India working on various areas like middleware and cloud computing. He then moved to Yahoo engaging academia on solving problems relevant to industry leveraging research in ML, IR, NLP and data mining. He's passionate about catalyzing industry−relevant data science in global universities.

He has published articles in conferences like IEEE SRDS as well as journals like TKDE. Chelliah presented following tutorial:

- Chelliah M, Sarkar S. : Product Recommendations Enhanced with Reviews. RecSys-2017.
- Concept to Code: Neural Network for Sequence Learning. ECIR 2019 (Full day tutorial Accepted)
- Personalised Fashion Recommendation Using Deep Learning, CoDS/ COMAD. Jan 2019
- Concept to Code: Deep Learning for Fashion Recommendation. The Web Conference 21019 (Accepted)

Asif Ekbal:
Email: asif@iitp.ac.in
Address: IIT Patna, India

Asif Ekbal is currently an Associate Professor in the Department of Computer Science and Engineering, IIT Patna. He has been pursuing research in Natural Language Processing (NLP), Information Extraction, Text Mining and Machine Learning (ML) applications for the last 11 years, and has made significant contributions in these areas. He has been handling several sponsored research projects in the broad areas of Natural Language Processing, Artificial Intelligence and Machine Learning technologies, funded by the different industries, such as Elsevier, Accenture and Samsung. He has co-authored around 150 papers in well-known foras such as ACM Transaction, ACL, COLING, EMNLP, HLT-NAACL, EACL etc. Google Scholar Citation which is the benchmark for Computer Science and Engineering shows 2297 career citations to his papers with h-index: 26 and i-10 index: 68. He is currently serving as senior program committee member for NAACL-2019 and IJCAI-22019. He has also served in the PC of several other conferences such as COLING, EMNLP, FLAIRS, ICON etc.

Mohit Gupta:
Email: mohit.gupta@flipkart.com
Address: Flipkart, Bangalore, India

Mohit has been part of Flipkart seller platform team for the last 2 years focused on review analysis for product insight extraction as a software development engineer. He was with Infosys earlier and holds an undergraduate degree in Computer Science from BITS, Pilani.