

Suggestion Mining from Opinionated Text

Sapna Negi

Insight Centre for Data Analytics
National University of Ireland, Galway
sapna.negi@insight-centre.org

Abstract

In addition to the positive and negative sentiments expressed by speakers, opinions on the web also convey suggestions. Such text comprise of advice, recommendations and tips on a variety of points of interest. We propose that suggestions can be extracted from the available opinionated text and put to several use cases. The problem has been identified only recently as a viable task, and there is a lot of scope for research in the direction of problem definition, datasets, and methods. From an abstract view, standard algorithms for tasks like sentence classification and keyphrase extraction appear to be usable for suggestion mining. However, initial experiments reveal that there is a need for new methods, or variations in the existing ones for addressing the problem specific challenges. We present a research proposal which divides the problem into three main research questions; we walk through them, presenting our analysis, results, and future directions.

1 Introduction

Online text is becoming an increasingly popular source to acquire public opinions towards entities like persons, products, services, brands, events, social debates etc. State of the art opinion mining systems primarily utilise this plethora of opinions to provide summary of positive and negative sentiments towards entities or topics. We stress that opinions also encompass suggestions, tips, and advice, which are often explicitly sought by stakeholders. We collaboratively refer to this kind of information as suggestions. Suggestions about a variety of topics of interest may be found on opin-

ion platforms like reviews, blogs, social media, and discussion forums. These suggestions, once detected and extracted, could be exploited in numerous ways. In the case of commercial entities, suggestions present among the reviews can convey ideas for improvements to the brand owners, or tips and advice to customers.

Suggestion extraction can also be employed for the summarisation of dedicated suggestion forums¹. People often provide the context in such posts, which gets repetitive over a large number of posts. Suggestion mining methods can identify the exact textual unit in the post where a suggestion is conveyed.

Table 1 provides examples of suggestions found in opinion mining datasets. In our previous work (Negi and Buitelaar, 2015b), we showed that suggestions do not always possess a particular sentiment polarity. Thus the detection of suggestions in the text goes beyond the scope of sentiment polarity detection, while complements its use cases at the same time.

In the recent past, suggestions have gained the attention of the research community. However, most of the related work so far performs a binary classification of sentences into *suggestions* or *non-suggestions*, where suggestions are defined as the sentences which propose improvements in a reviewed entity (Brun and Hagege, 2013; Ramanand et al., 2010; Dong et al., 2013). These studies annotated datasets accordingly and developed systems for the detection of only these type of suggestions; and performed an in-domain evaluation of the classifier models on these datasets.

We emphasise that in addition to the classification tasks performed earlier, there are a lot more aspects associated with the problem, including a well-formed and consistent problem defini-

¹<https://feedly.uservoice.com/forums/192636-suggestions/category/64071-mobile>

tion. We divide the study of suggestion mining into three guiding aspects or research questions:

1) Definition of suggestions in the context of suggestion mining, 2) Their automatic detection from opinionated text, and 3) Their representation and summarisation.

A comprehensive research on suggestion mining demands the problem specific adaptation and integration of common NLP tasks, like text classification, keyphrase extraction, sequence labelling, text similarity etc. Last but not least, recent progress in the adaptation of deep learning based methods for NLP tasks opens up various possibilities to employ them for suggestion mining.

2 Research Problem

A broad statement of our research problem would be, *mining expressions of suggestions from opinionated text*. There are several aspects of the problem which can lead to a number of research questions. We identify three broad research questions which are the guiding map for our PhD research.

- Research Question 1 (RQ1): How do we define suggestions in suggestion mining?
- Research Question 2 (RQ2): How do we detect suggestions in a given text ?
- Research Question 3 (RQ3): How can suggestions be represented and summarised ?

The following sections will give a more detailed description of these aspects, including the preliminary results, challenges, and future directions.

3 Research Methodology

In this section we address each of the research questions, our findings so far, and the future directions.

3.1 RQ1: Suggestion Definition

The first sense of *suggestion* as listed in the oxford dictionary is, *an idea or plan put forward for consideration*, and the listed synonyms are *proposal, proposition, recommendation, advice, counsel, hint, tip, clue* etc. This definition, however needs to be defined on a more fine grained level, in order to perform manual and automatic labelling of a text as an expression of suggestion.

There have been variations in the definition of suggestions targeted by the related works, which renders the system performances from

some of the works incomparable to the others. We identify three parameters which can lead us to a well-formed task definition of *suggestions* for suggestion mining task: What is the unit of a suggestion, who is the intended receiver, and whether the suggestion is expressed explicitly or not.

Unit: Currently, we consider sentence as a unit of suggestion, which is in-line with related works. However, it was observed that some sentences tend to be very long, where suggestion markers are present in only one of the constituent clauses. For example: *When we booked the room the description on the website said it came with a separate seating area, despite raising the issue with reception we were basically told this was not so , I guess someone needs to amend the website*. In this sentence, although the full sentence provides context, the suggestion is identifiable from the last clause. It is common to witness such non-uniform choice of punctuation in online content. Considering this, we intend to build classification models which can identify the exact clause/phrase where a suggestion is expressed, despite of individual instances being sentences.

Receiver: Different applications of suggestion mining may target different kinds of suggestions, which can differ on the basis of intended receiver. For example, in domains like online reviews, there are two types of intended receivers, brand owners, and fellow customers. Therefore, suggestions need to be defined on the basis of the intended receivers.

How is a suggestion expressed: The first round of suggestion labelling performed by us resulted in a very low inter-annotator agreement, i.e. a kappa score of 0.4 - 0.5. It was observed that given a layman definition of suggestions, humans do not distinguish between explicit and implicit forms of suggestions, since they can inherently infer suggestions from their implicit forms. Figure 1 illustrates the two forms. Specifically, in the case of domains like reviews, annotators mostly disagreed on whether the implicit ones are suggestions or not. We define an explicit suggestion as the text which directly proposes, recommends, or advices an action or an entity; whereas the implicit ones provide the information

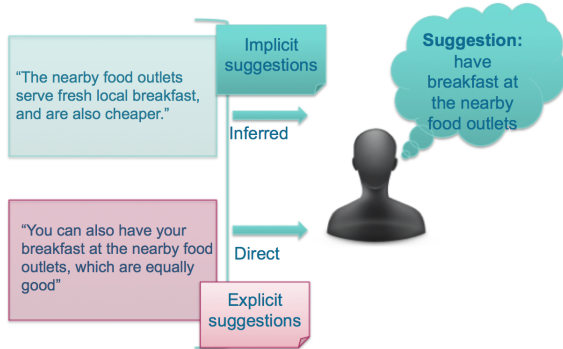


Figure 1: Implicit and explicit forms of suggestions

from which the suggested action or entity can be inferred. In remainder of the paper, we refer to explicit suggestions as *suggestions*.

We observe that certain linguistic properties consistently mark suggestions across different datasets (Table 1). One such phenomenon is imperative and subjunctive mood (Negi and Buitelaar, 2015a; Negi and Buitelaar, 2015b). The presence of these properties makes it more likely, but does not guarantee a text to be a suggestion. Another linguistic property is *speech act* (Searle, 1969). Speech act is a well studied area of computational linguistics, and several typologies for speech acts exist in literature, some of which consider suggestions as a speech act (Zhang et al., 2011).

3.2 RQ2: Suggestion Detection

The problem of suggestion detection in a big dataset of opinions can be defined as a sentence classification problem: Given a set S of sentences $\{s_1, s_2, s_3, \dots, s_n\}$, predict a label l_i for each sentence in S , where $l_i \in \{\text{suggestion, non suggestion}\}$.

The task of suggestion detection rests on the hypothesis that a large amount of opinionated text about a given entity or topic is likely to contain suggestions which could be useful to the stakeholders for that entity or topic. This hypothesis has been proven to be true when sentences from reviews and tweets about commercial entities were manually labeled (Table 1). Also, the survey presented by Asher et al. (2009) shows that although in a low proportion, opinionated texts do contain expressions of advice and recommendations.

The required datasets for suggestion based sentence classification task are a set of sentences which are labelled as *suggestion* and *non-suggestion*, where the labeled suggestions should be explicitly expressed.

Existing Datasets: Some datasets on suggestions for product improvement are unavailable due to their industrial ownership. To the best of our knowledge, only the below mentioned datasets are publicly available from the previous studies:

- 1) Tweet dataset about Microsoft phones: comprises of labeled tweets which give suggestions about product improvement (Dong et al., 2013). Due to the short nature of tweets, suggestions are labeled at the tweet level, rather than the sentence level.
- 2) Travel advice dataset: comprises of sentences from discussion threads labeled as *advice* (Wicaksono and Myaeng, 2013). We observe that the statements of facts (implicit suggestions/advice) are also tagged as advice in this dataset, for example, *The temperature may reach upto 40 degrees in summer*. Therefore, we re-labeled the dataset with the annotation guidelines for explicit suggestions, which reduced the number of positive instances from 2192 to 1314.

Table 2 lists the statistics of these datasets.

Introduced Datasets: In our previous work (Negi and Buitelaar, 2015b), we prepared two datasets from hotel and electronics reviews (Table 2) where suggestions targeted to the fellow customers are labeled. Similar to the existing Microsoft tweets dataset, the number of suggestions are very low in these datasets. As stated previously, we also formulate annotation guidelines for the explicit expression of suggestions, which led to a kappa score of upto 0.86 as the inter-annotator agreement. In another work (Negi et al., 2016), we further identify possible domains and collection methods, which are likely to provide suggestion rich datasets for training statistical classifiers.

- 1) Customer posts from a publicly accessible suggestion forums for the products *Feedly mobile app*², and *Windows App studio*³. We crawled

²<https://feedly.uservoice.com/forums/192636-suggestions>

³<https://wpdev.uservoice.com/forums/110705-universal-windows-platform>

Source, Entity/Topic	En-	Sentence	Intended Receiver	Sentiment	Linguistic Properties
Reviews, Electronics		I would recommend doing the upgrade to be sure you have the best chance at trouble free operation.	Customer	Neutral	Subjunctive, Imperative, lexical clue: <i>recommend</i>
Reviews, Electronics		My one recommendation to creative is to get some marketing people to work on the names of these things	Brand owner	Negative	Imperative, lexical clue: <i>recommendation</i>
Reviews, Hotels		Be sure to specify a room at the back of the hotel.	Customer	Neutral	Imperative
Tweets, Windows Phone		Dear Microsoft, release a new zune with your wp7 launch on the 11th. It would be smart	Brand owner	Neutral	Imperative, subjunctive
Discussion thread, Travel		If you do book your own airfare, be sure you don't have problems if Insight has to cancel the tour or reschedule it	Thread participants	Neutral	Conditional, imperative

Table 1: Examples of similar linguistic properties in suggestions from different domains, about different entities and topics, and intended for different receivers

the suggestion for improvement posts for these products, and labeled only a subset of them due to the annotation costs. Although all the posts are about suggestions, they also comprise of explanatory and informative sentences around the suggestion sentences. With the availability of more annotation resources, this dataset can be easily extended.

2) We also prepared a new tweet dataset, where the tweets are first collected using the hashtags *suggestion*, *advice*, *recommendation*, *warning*, which appeared as top unigram features in our SVM based classification experiments. This sampling method increased the likelihood of the presence of suggestion tweets as compared to the Microsoft tweets dataset.

Table 2 details all the currently available datasets including the old and the new ones.

Sentence Classification: Conventional text classification approaches, including, rule based classifiers, and SVM based classifiers have been previously used for this task. We employ these two approaches on all the available datasets as baselines. In addition to the in-domain training and evaluation of statistical classifiers, we also perform a cross-domain training and evaluation. The reason for performing a cross domain training experiment is that the suggestions possess similar linguistic properties irrespective of the domain (Table 1). Since, it is expensive to prepare dedicated training dataset for each domain or use case, we aim for domain independent classification models.

We performed a first of its kind study of the employability of neural network architectures like Long Short Term Memory (LSTM), and Convolutional Neural Nets (CNN) for suggestion detection. The F-scores for positive class are shown in Table 2. A neural network based approach seems to be promising compared to the baseline

approaches, specifically in the case of domain independent training. Our intuition is that the ability of word embeddings to capture semantic and syntactic knowledge, as well as the ability of LSTM to capture word dependencies are the contributing factors to this.

There is a lot of scope for improvement in the current results. One challenge is that the sentences are often longer, whereas the suggestion is present only as a phrase or clause. Therefore, a future direction is to explore sequential classification approaches in this regard, where we can tag sentences at the word level, and train the classifiers to predict binary labels corresponding to whether a word is a part of suggestion or not. For example, *My_1 recommendation_1 is_1 to_1 wait_1 on_1 buying_1 one_1 from_1 this_1 company_1 as_0 they_0 will_0 surely_0 get_0 sent_0 a_0 message_0 of_0 many_0 returned_0 dvd_0 players_0 after_0 christmas_0*. LSTM NNs have also been proven to be a good choice for sequence labelling tasks (Huang et al., 2015).

3.3 Suggestion Representation and Summarisation

In order to apply suggestion mining to real life applications, a more structured representation of suggestions might be required. After the extraction of suggestion sentences from large datasets, there should be a way to cluster suggestions, link them to relevant topics and entities, and summarise them. One way of achieving this is to further extract information from these sentences, as shown in Table 3.

We start with the task of extracting the central phrase from a suggestion, which either corresponds to a recommended entity or a suggested action. As a first step in this direction, we experimented with keyphrase extraction. Keyphrase extraction has been mainly used for the detection of topical information, and is therefore noun-based

Dataset	Intended receiver	No. of suggestions	F1 score			
			Rules	SVM	LSTM	CNN
In-domain Evaluation						
Hotel Reviews	Customers	448 / 7534	0.285	0.543	0.639	0.578
Electronics reviews	Customers	324 / 3782	0.340	0.640	0.672	0.612
Travel advice	Thread participants	1314 / 5183	0.342	0.566	0.617	0.586
Tweets (Microsoft)	Brand owner	238 / 3000	0.325	0.616	0.550	0.441
New Tweets	Public	1126 / 4099	0.266	0.632	0.645	0.661
Suggestion Forum	Brand owners	1428 / 5724	0.605	0.712	0.727	0.713
Cross-domain Evaluation						
Training Dataset	Test Dataset	No. of suggestions (training)	F1 score			
			Rules	SVM	LSTM	CNN
Sugg-Forum	Hotel	1428 / 5724	0.285	0.211	0.452	0.363
Sugg-Forum	Electronics	1428 / 5724	0.340	0.180	0.516	0.393
Sugg-Forum	Travel advice	1428 / 5724	0.342	0.273	0.323	0.453
Sugg-Forum + Travel advice	Hotel	2742 / 10907	0.285	0.306	0.345	0.393
Sugg-Forum + Travel advice	Electronics	2742 / 10907	0.340	0.259	0.503	0.456
New Tweets	Microsoft Tweets	1126 / 4099	0.325	0.117	0.161	0.122

Table 2: Results of suggestion detection across datasets, using different methods

Full suggestion text	Entity	Beneficiary	Keyphrase
If you do end up here, be sure to specify a room at the back of the hotel	Room	Customer	Specify a room at the back of the hotel
If you are here, I recommend a Trabi safari	Trabi Safari	Customer	Trabi Safari
Chair upholstery seriously needs to be cleaned	Chair/Chair upholstery	Brand owner	chair upholstery need to be cleaned

Table 3: Aimed information extraction from suggestions

(Hasan and Ng, 2014). As Table 3 shows, we also need to detect verb based keyphrases in the case of advice or action based suggestions, however a noun based keyphrase would work in the case of suggestions which recommend an entity.

In the Table 4, we show the examples of keyphrases extracted using *TextRank* (Mihalcea and Tarau, 2004) algorithm on 3 different review datasets, i.e. ebook reader, camera, and hotel. *TextRank* and almost all of the keyphrase extraction algorithms rely on the occurrence and co-occurrence of candidate keyphrases (noun phrases) in a given corpus. We ran *TextRank* on the reviews and obtained a set of keyphrase. Table 4 shows whether the central phrases contained in a suggestion from the dataset were detected as a keyphrase by the algorithm or not. In the case of suggestion for improvement i.e. sentence 1, *TextRank* is able to capture relevant noun keyphrases. This can be attributed to a large number of sentences in the corpus which mention *price*, which is an important aspect of the reviewed entity. However, in the case of suggestions which are addressed to the other customers, reviewers often speak about aspects which do not appear frequently in reviews. This can be observed in sentence 2 and 3, where the keyphrase were not de-

tected.

We plan to include keyphrase annotations to the sequence labels mentioned in section 3.2, in order to identify the suggestions as well as the keyphrases within those suggestions at the same time.

After the representation of suggestions in the proposed format, we plan to use the methods for text similarity and relatedness in order to cluster similar suggestions.

Suggestions	Extracted keyphrase	Desired keyphrase
Look around and compare price...the price of the Nook varies widely between stores.	Price	compare price
I suggest that you purchase additional memory	<i>none</i>	purchase additional memory
Would recommend the Kohi Noor indian around the corner, book in advance as it is small and well used by locals.	<i>none</i>	Kohi Noor Indian, book in advance

Table 4: Sample results from keyphrase extraction using the *textRank* algorithm

4 Related Work

To the best of our knowledge, all of the related work so far focussed on research question 2, i.e. detection of suggestions in a given text. In recent years, only a limited number of experiments have

been performed in this regard, all of which frame the task as a sentence classification task.

Suggestion Detection: Ramanand et al. (2010) and Brun et al. (2013) employed manually crafted linguistic rules to identify suggestions for product improvement. Dong et al. (2013) performed classification of given tweets about Microsoft Windows' phone as suggestions for improvement or not. Wicaksono et al. (Wicaksono and Myaeng, 2013) detected advice containing sentences from travel related discussion threads. They employed sequential classifiers, based on Hidden Markov Model, and Conditional Random Fields. Most of the datasets are not available, and annotations for the available datasets are ambiguous with no detailed definition of what is considered as a *suggestion* or *advice*.

Text Classification using deep learning: Recent advances in the application of Neural Networks (NNs) to NLP tasks demonstrate their effectiveness in some of the text classification problems, like sentiment classification, pos tagging, and semantic categorisation. Long Short Term Memory NNs (Graves, 2012), and Convolutional NNs (Kim, 2014) are the two most popular neural network architectures in this regard. An end to end combination of CNN and LSTM (Zhou et al., 2015) has also shown improved results for sentiment analysis.

5 Conclusion

In this work we presented a research plan on suggestion mining. The problem in itself introduces a novel information mining task. Several useful datasets have already been released, with more to come. The related work in this direction is very limited, and has so far focussed on only one aspect of the problem. Our proposal proposes research contributions in three research aspects/questions, and presents initial results and analysis.

Since suggestions tend to exhibit similar linguistic structure, irrespective of topics and intended receiver of the suggestions, there is a scope of learning domain independent models for suggestion detection. Therefore, we test the discussed approaches both in a domain-independent setting as well, in order to test the domain-independence of models learnt in these approaches. Neural networks in general outperformed the results on existing test datasets, in both domain dependent and independent training. In light of these findings,

building neural network based classification architectures for intra-domain feature learning can be an interesting future direction for us.

The results also point towards the challenges and complexity of the task of suggestion mining. Building word level suggestion tagged datasets seems to be a promising direction in this regard, which can simultaneously address the tasks of suggestion detection and as keyphrase extraction for suggestion mining.

Our research findings and datasets can also be employed to similar problems, like classification of speech acts, summarisation, verb based keyphrase extraction, and cross domain classification model learning.

Acknowledgements

This work has been funded by the European Unions Horizon 2020 programme under grant agreement No 644632 MixedEmotions, and the Science Foundation Ireland under Grant Number SFI/12/RC/2289 (Insight Center).

References

- [Asher et al.2009] Nicholas Asher, Farah Benamara, and Yannick Mathieu. 2009. Appraisal of Opinion Expressions in Discourse. *Linguistic Investigations*, 31.2:279–292.
- [Brun and Hagege2013] C. Brun and C. Hagege. 2013. Suggestion mining: Detecting suggestions for improvements in users comments. In *Proceedings of 14th International Conference on Intelligent Text Processing and Computational Linguistics*.
- [Dong et al.2013] Li Dong, Furu Wei, Yajuan Duan, Xiaohua Liu, Ming Zhou, and Ke Xu. 2013. The automated acquisition of suggestions from tweets. In Marie desJardins and Michael L. Littman, editors, *AAAI*. AAAI Press.
- [Graves2012] Alex Graves. 2012. Supervised sequence labelling. In *Supervised Sequence Labelling with Recurrent Neural Networks*, pages 5–13. Springer Berlin Heidelberg.
- [Hasan and Ng2014] Kazi Saidul Hasan and Vincent Ng. 2014. Automatic keyphrase extraction: A survey of the state of the art. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1262–1273, Baltimore, Maryland, June. Association for Computational Linguistics.
- [Huang et al.2015] Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional LSTM-CRF models for sequence tagging. *CoRR*, abs/1508.01991.

- [Kim2014] Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.
- [Mihalcea and Tarau2004] R. Mihalcea and P. Tarau. 2004. TextRank: Bringing order into texts. In *Proceedings of EMNLP-04 and the 2004 Conference on Empirical Methods in Natural Language Processing*, July.
- [Negi and Buitelaar2015a] Sapna Negi and Paul Buitelaar. 2015a. Curse or boon? presence of subjunctive mood in opinionated text. In *Proceedings of the 11th International Conference on Computational Semantics*, pages 101–106, London, UK, April. Association for Computational Linguistics.
- [Negi and Buitelaar2015b] Sapna Negi and Paul Buitelaar. 2015b. Towards the extraction of customer-to-customer suggestions from reviews. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2159–2167, Lisbon, Portugal, September. Association for Computational Linguistics.
- [Negi et al.2016] Sapna Negi, Kartik Asooja, Shubham Mehrotra, and Paul Buitelaar. 2016. A distant supervision approach to semantic role labeling. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, Berlin, Germany, August. Association for Computational Linguistics.
- [Ramanand et al.2010] J Ramanand, Krishna Bhavsar, and Niranjana Pedanekar. 2010. Wishful thinking - finding suggestions and 'buy' wishes from product reviews. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 54–61, Los Angeles, CA, June. Association for Computational Linguistics.
- [Searle1969] John R. Searle. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, London.
- [Wicaksono and Myaeng2013] Alfan Farizki Wicaksono and Sung-Hyon Myaeng. 2013. Automatic extraction of advice-revealing sentences for advice mining from online forums. In *K-CAP*, pages 97–104. ACM.
- [Zhang et al.2011] Renxian Zhang, Dehong Gao, and Wenjie Li. 2011. What are tweeters doing: Recognizing speech acts in twitter. In *Analyzing Microtext*, volume WS-11-05 of *AAAI Workshops*. AAAI.
- [Zhou et al.2015] Chunting Zhou, Chonglin Sun, Zhiyuan Liu, and Francis C. M. Lau. 2015. A C-LSTM neural network for text classification. *CoRR*, abs/1511.08630.