Exploiting the Focus of the Document for Enhanced Entities' Sentiment Relevance Detection

Zvi Ben-Ami
The Hebrew University
zvi.benami@mail.huji.ac.il

Ronen Feldman
The Hebrew University
ronen.feldman@huji.ac.il

Benjamin Rosenfeld
Amenity Analytics
grurgrur@gmail.com

Abstract— **A key question in sentiment analysis is whether sentiment expressions, in a given text, are related to particular entities. This is an imperative question, since people are typically interested in sentiments on specific entities and not in the overall sentiment articulated in an article or a document. Sentiment relevance is aimed at addressing this precise problem. In this paper, we argue that exploiting information about the focus of the document on the entity of interest can significantly improve the task of detecting sentiment relevance and, hence, the final sentiment scores assigned for the entities. In order to assess the value of such information, we look at various methods for detecting sentiment relevance for entities. We consider both rule-based algorithms that rely on the entity's physical or syntactic proximity to the sentiment expressions as well as more sophisticated machine learning classification algorithms. We demonstrate that the focus of the document on the entities within it is, indeed, an important piece of information, which can be accurately learned with supervised classification means. We, further, found that overall classification-based algorithms perform better than the deterministic ones in identifying sentiment relevance, with sequence-classification performing significantly better than direct classification.**

Keywords— Sentiment Analysis, Sentiment Relevance, Document-level Information, Focus of the Document, Document Type with Respect to Entity, Entity-level Sentiment Analysis

I. INTRODUCTION

Sentiment Analysis (SA) is driven by the desire to understand the explicit thoughts and opinions of people on particular entities or situations, in a given text [1], [2]. A requisite task is to detect, in the text, both the entity of interest and sentiment expressions. Effectively identifying both, however, does not ensure that the particular expression truly pertains to the entity of interest. In order to ensure that, one should bind the sentiment expressions with the underlining entity, or in other words, determine the sentiment relevance for the entity of interest. This level of granularity in SA is often referred to in the literature as entity-level SA.

To address this problem, researchers suggested various rule-based methods that use domain-specific lexicons (e.g.,[3]). Other researchers proposed applying Machine Learning (ML) techniques (e.g., [4]), typically classification with supervised or semi-supervised means, which may exploit diverse sets of features representations (e.g., [5], [6]).

In this paper, we argue that a better binding of sentiment expressions with entities can be achieved by considering context information about the entity of interest, for example, the type of the entity, whether it is the only entity of its type in the document or one of many, the position of the entity within the document, section, or paragraph, the lexical contents, etc. We demonstrate that one of the most valuable context clues is the "focus of the document on the entity of interest", i.e., whether the entity is the main topic of the document being analyzed, one of several main topics, or is just mentioned in passing. We further illustrate that this information can be successfully obtained with supervised ML means.

One possible strategy for entity-level SA would be to first identify sentiments expressions in the text and then decide to which entities they refer. An alternative strategy would be to first identify the passages within the text that are relevant to each entity of interest and then analyze the sentiments within them. The choice of which entity to bind a sentiment expression to can then be made according to the physical, syntactical, and/or semantic proximity to the entities of interest and according to the salience of the entities in the documents. We establish that all of these methods can be useful in different scenarios and, therefore, the best single algorithm should use all available proximity information, of all kinds, together with additional context information.

Understanding the nature of particular entity types and the way they interact with entities of other types are additional properties that should be considered when determining the sentiment relevance for entities. In this paper, we distinguish between cases where the interest is in one entity type, such as a company or a product, and cases where the interest is in the interaction between entities of different types, for example the sentiment on a certain drug when treating a certain disease.

In this study, we pay particular attention to two domains of interest: Financial and Medical, but the results can be expanded to other domains. The main contributions of this study are on three levels, first, we show that sentiment relevance detection is important for the general task of sentiment analysis. Second, we show that the focus of the document is valuable information that should be considered when determining sentiment relevance, and that it can be learned with supervised ML means. Third, we compare between the performances of various methods for detecting sentiment relevance.

978-1-4673-8493-3/15 $31.00 © 2015 IEEE
DOI 10.1109/ICDMW.2015.37

The remainder of this paper is structured as follows: in the next section, we review the related work. In Section III, we discuss in greater detail the problem of entity sentiment relevance and distinguish between cases where one is interested in sentiment relevance for a single entity-type vs. multiple entity-type. In Section IV, we describe the different relevance detection algorithms that are considered in this paper. Our experiments and experiments results are, then, presented in Section V. In the last section, we conclude our study and propose future work.

II. RELATED WORK

The task of SA has drawn the attention of many researchers around the world for over a decade [1], [2], [7]–[10]. While most SA research is focused on discovering and classifying the sentiment expressions, some studies are also concerned with the targets of the expressions [11].

Existing sentiment analysis systems are typically differentiated on two general dimensions. The first is the level of granularity at which the analysis is being conducted and the second is the level of automatization, i.e., the extent to which ML versus rule-based techniques are applied. In this work, we primarily study the effect of information on the focus of the document in determining sentiment relevance and, therefore, look at a wide spectrum of methods for SA.

With regards to the level of granularity, simplified SA systems approach the problem on a document-level basis [12], [13]. The underlining assumption of this level of analysis is that the document contains an opinion on one main object expressed by its author [1]. This level of analysis is valuable if one wishes to discover the overall tone of an article or if the article is indeed on one specific topic. For instance, in movie reviews, each review is typically on one particular movie and, therefore, this level of granularity is sufficient if the objective is to identify the general sentiment toward it.

Documents, however, tend to discuss more than one entity and, since people are typically interested in sentiments on specific entities and not in the overall sentiment articulated in a document, more advanced approaches for SA attempt to address this problem at a finer level of analysis. In that sense, a sentence-level analysis is more logical, as it is more likely that the sentence discusses only one entity. This approach is, therefore, rather popular [14]–[18], however, it tends to suffer from two main pitfalls: the first is that even in a single sentence more than one entity (of similar or different types) could be mentioned, and the second is that a sentence without explicit reference to the entity of interest could still be relevant to the particular entity, even if all coreferences were correctly resolve in the text. When analyzing financial articles, for instance, sentences with sentiment that does not mention the company of interest (directly or indirectly), may still be relevant to the company if they discuss its' CEO, one of its' products or its recent annual reports. In such cases, a sentence-level analysis will suffer from many recall errors.

Some SA systems try to take even finer-grain approaches and look into syntactic or physical proximity in order to better link between the entities of interest and existing sentiment expressions. Hu and Liu [15], for example, introduced a set of techniques for a feature-based summary of online customer reviews. In order to associate sentiments expressions with features of entities, they used a simplified approach of nearest noun/noun phrases. Pand and Lee [8] proposed features-based opinion mining by explicitly identify the syntactic targets of sentiment expressions as features or aspects of the target entity. Jiang et al. [5] used several target-dependent and target-independent features for sentiment classification of tweets, the target-dependent features were generated using a syntactic parse tree and a set of rules.

In the domain of product reviews, a common approach is to start with identifying the features (characteristics) of the products and then create a lexicon of sentiment expressions related to each feature [3], [19], where the association of sentiment expression with the particular features is typically done with rule-based approaches.

Other related works belong to the Passage Retrieval field, since the sentiment relevance detection problem can be constructed as a specific form of a passage retrieval problem. Passage retrieval pertains to the task of identifying relevant pieces or passages of information within an unstructured text document that discusses several subject areas, topics or entities [20], [21]. In many cases, passages can be recognizable units, such as paragraphs and sentences [22]–[24]. Because of the aforementioned problem related to this level of analysis, early studies suggested the use of a window of text around entities of interest. The windows consist of a fixed number of words or bytes, which might be overlapping, disjointed or partly overlapping and may or may not rely on the logical structure of the document [20], [25]. More recent papers (e.g., [26], [27]), proposed more sophisticated approaches, for example, to use n-gram or dynamically adjust the number of keywords used for retrieval. Wachsmuth [28] suggested using an input control that maintains the dependencies between all relevant types of entities and relations in order to analyze and filter only relevant portions of text.

While some SA systems exploit such rules to establish this link, others approached the problem using ML classification techniques (e.g., [12]), in which sentiment expressions are typically classified as relevant or irrelevant to an entity using annotated training data. Jakob and Gurevych [6] used a set of features to train a Conditional Random Fields (CRF) classifier for Opinion Target Extraction. The features used in their model include tokens, POS, dependency path, word distance and opinion sentences. Engonopoulos et al. [4] suggested a method that splits sentences into smaller segments, each of which refers at most to one entity, and conveys a single sentiment (or none) towards it. They performed an entity-level sentiment classification that is based on sequential modeling using CRF and demonstrated that this approach produces better results than rule-based methods, such as those presented in [3], [29].

Zhang et al. [16] suggested a hybrid method that first uses a augmented lexicon-based method for entity-level to generate training data for a binary sentiment classifier that assigns sentiment polarities to the newly-identified opinionated tweets. Since the level of analysis is sentence, their underlining assumption is that sentiment expressions are necessarily associated with the entities mentioned in the same sentence.

Scheible and Schütze [17] defined the "S-relevance" concept and investigated some of its properties. In contrary to our work, they focused on distinguishing sentences which are S-relevant, i.e., possess informative content for determining the sentiment, and sentences which are not S-relevant. They found this distinction to be more appropriate for sentiment analysis than the well-established subjective/objective categorization. They investigated two semi-supervised approaches to relevance classification, namely, a distant supervision approach that exploits structured information about the domain; and transfer learning on feature representations based on lexical taxonomies. In contrast to Scheible and Schütze [17], we strive to discover sentiments relevance for all entities (of a given type) mentioned in the document, not necessarily topical.

To the best of our knowledge, no other work on entity-level SA studied and exploited the context information on the focus of the document on the entity in detecting sentiment relevance. Our work is also novel as we distinguish between cases where we are interested in single entity-type sentiment relevance and multiple entity-type relevance detection problems (See Section III) and compare between the accuracy of various proximity-based and classification-based algorithms (See section IV) in the problem of sentiment relevance.

III. Sentiment Relevance for Different Entity Types

The nature of particular entity types and the way they interact with entities of other types are properties that should be considered when determining the sentiments relevancy of entities. In many cases, we may only be interested in the sentiments on one entity type. For example, when analyzing financial news articles, we are likely to be interested in sentiments about companies, so our target entity type would be COMPANY; when analyzing political news we may be interested in sentiments about political candidates, so our target entity type would be PERSON; when analyzing product reviews, we will look for sentiments about the products and our target entity type would be PRODUCT, etc. In such cases, clearly distinguishing between the relevancy of passages related to the target and to the non-target entities types is not essential. For instance, in the financial domain, when the general topic is a COMPANY, and there is a sentiment expression referring to a PERSON, such as the CEO of the company, or a PRODUCT of the company, this sentiment expression is still relevant to the company and can be regarded as such during the sentiment analysis, as if the entities of other types were not present at all. Sometimes, for the purposes of sentiment analysis, such semantically related entities of different entity types may themselves be considered features (or aspects) of the target entity [1], [30]. Aspect-based sentiment analysis is, however, a different topic from sentiment relevance discussed in this paper. In this regard, we are interested in how the different types of entities should be handled when determining sentiment relevance, rather than on analyzing the sentiments on features or aspects of a particular entity type.

In other situations, people may not be interested in only one entity type, but rather in the specific interaction between entities of different types. For example, when analyzing medical forums, we may want to answer questions such as: “Do people think this drug is good for curing this disease?”, so we are essentially interested in the sentiments on a given DRUG in the context of a given DISEASE or i.e., the interaction between DRUG and DISEASE. We can also think of an additional interaction that involves a particular PERSON, to answer questions such as: "Does this expert physician think this drug is good for curing this disease?”. We will show that such situations are modeled well enough using intersections of regions of relevance of the corresponding entity types, while calculating the relevance region for each type separately.

We purposefully exclude possible interactions between entities of the same type, such as interactions of sentiments regarding two companies or two people, because they behave in a distinct way compared to the interactions between entities of different types. In terms of sentiment relevance, the ranges of the interacting same-type entities simply overlap, and the sentiment expressions in the common section belong to both entities. The interactions themselves may be conjunctive or adversative, with the sentiment polarity being the same or flipping to one of the entities. The precise analysis of such interactions falls outside the scope of the relevance detection problem, and so it is mostly ignored in this paper.

A. Single-entity Sentiment Relevance

A single-entity sentiment relevance detection problem instance consists of the triple: “document”, “sentiment expression within the document” and “target entity instance”. Our goal is to determine whether the sentiment expression within the document is relevant to the target entity instance. Hence, the task can be defined as the binary decision: 'relevant' vs. 'irrelevant'. To address this problem, we can use any information that can be found by analyzing the document. Thus, we can assume to know the structure of the document (paragraphs and sentences boundaries), the parse trees of all sentences, as well as the locations of all references of all entities in the document, including coreferences.

In this work, we suggest making use of an extra piece of information regarding each target entity, which is the "focus of the document on the entity of interest", or similarly from a different perspective the "document type with respect to the entity". We distinguish between four intuitive types, which are clearly different:

- 'Main_Topic' – the entity is the main topic in the document, i.e. the document focuses on the entity of interest;
- ‘Minor_Topic’ – the entity is not the main topic in the document, and is mentioned in passing, i.e., the document focuses on another entity;
- 'Related_Topics' – the main topic in the document is a relation between the entity and some other entities of the same type, i.e., the document focuses on some connections between two or more entities of the same type; and
- 'Few_Topics' – the entity is one of a few equally important topics, dealt with sequentially, i.e., the

document focuses on several entities of the same type which are being discussed separately in the document.

The same document would, therefore, be of different types with respect to different entities mentioned within it. In the datasets we use for the experiments, the document type is manually annotated with respect to each entity, this allows us to directly observe the influence of this information on detecting sentiment relevance. We show that this information can also be automatically extracted using a supervised classification learning method by using part of the annotated data as training data for the classifier (see section V-B).

B. Multiple-entity-types Relevance

When a sentiment expression is relevant to several entities of different types, we encounter a problem of multiple-entity-types relevance, which is distinguished from the single-entity relevance problem described above as the concern is about entities of several types rather than about only one type.

This problem is close to Relation Extraction in a sense. In particular, we are interested in examples from the medical domain regarding three entity types: PERSON, DRUG, and DISEASE, where PERSON is restricted to known physicians. While each of the entity types can be the target of a sentiment expression, the more interesting questions in this domain involve multiple entities, specifically, the intersection of DRUG and DISEASE, in order to address questions such as "how effective is this drug for this disease?", and PERSON + DRUG + DISEASE, in order to address questions such as "what does this physician say about using this drug to cure this disease?". For illustration, consider the following text in Fig. 1, taken from the medical domain.

The paragraph in Fig. 1 is separately relevant for "Plegridy" among DRUGs, for "MS" (Multiple Sclerosis) among DISEASEs, and for "Peter Calabresi" among PERSONs, so all sentiment expressions ("positive effects", "offer the benefit", "meaningful advance") will be interpreted as multiply-entity-types relevant for these three entities.

Fig. 1. An extract from the medical corpus, with marked entities

> *"... Peter Calabresi, director of the Johns Hopkins Multiple Sclerosis Center said, "These full first-year results provide a more complete picture of Plegridy and its positive effects on the reduction of relapse, disability progression and lesion development. These data suggest that, if approved, Plegridy may offer the benefit of a less frequent dosing schedule, which would be a meaningful advance for people living with MS,", reports RTT news...."*

We solve the multiple-entity relevance problem by intersecting the relevance ranges of different-type entities, thus reducing the problem to the single-entity relevance detection. As such, the experiments regarding the multiple-entity relevance need only check the accuracy of this reduction. In the medical domain, at least, this accuracy appears to be adequate. The motivation for this is that when there is a sentiment expression together with a mentioning of both a DRUG and a DISEASE, it is obvious that the sentiment is with regards to the effect of the DRUG on the DISEASE and not vice-versa. Similarly, when a PERSON who is a known physician is mentioned together with a DRUG and a DISEASE, it is clear that the sentiment will reflect the physician's opinion on the drug's effect in treating the disease.

IV. RELEVANCE ALGORITHMS

To evaluate the influence of context information in SA models and, specifically, the focus of the document on the entity of interest, we look at a wide spectrum of SA algorithms, while providing or withholding this particular information. At the lower-end of complexity, we perform a document-level analysis. This simple model serves as a baseline in evaluating the overall necessity of detecting sentiment relevance. At the next level, we perform sentence-level analysis in which a sentiment expression is deemed relevant only if the entity of interest is explicitly mentioned (including co-reference) in the same sentence. At a finer granularity level, we look at algorithms of two types, rule-based algorithms that rely on the entity's proximity (physical and syntactic) to the sentiment expressions and ML algorithms (specifically, linear classification and sequence classification), which use a diversified set of features representation pertaining to the context in which the entity is mentioned. Using rule-based algorithms allow us to evaluate whether designing rules that are suited to the different possible foci of documents, can improve the performance of the detection task, which could, in turn, validate the benefit of exploiting this context information. The ML algorithms allow us to evaluate how well can this task be performed with supervised learning means.

In addition to evaluating various algorithms for detecting sentiment relevance with manually annotated data on the focus of the document, we evaluate these algorithms with annotated data that was obtained with supervised classification means. In other words, we first learned the focus of the document for each entity using supervised means, and then evaluated the performances of the different sentiment relevance algorithms using this automatically obtained data. Although many algorithms can be thought to be applicable for this task, comparing between the performances of different possible algorithms for identifying the focus of the document falls out of the scope of this study. The purpose of this particular exercise is to determine the feasibility of automatically identifying the focus of the document and not to find the most optimal methods.

In the following subsections, we describe in detail the particular algorithms used in our experiments. As a running example to illustrate the different algorithms, we use an extract from the financial corpus, shown in Fig. 2, which outlines a discussion on recent flash memory technologies and contains several sentiments on three companies: SanDisk, Samsung and Micron. The entity references (of the type COMPANY) are bold and the sentiment expressions are colored according to their polarity.

A. Document-level analysis

As a baseline, we construct a simple document-level sentiment analysis algorithm, in which every sentiment expression within the document is declared relevant for each mentioned entity.

Fig. 2. An extract from the financial corpus, with marked entities and sentiment expressions. (Expressions (1), (3)-(5), (12) are relevant for "SanDisk"; (2), (6)-(11) for Samsung; (5) for "Micron")

Flash memory maker **SanDisk (NASDAQ: SNDK)** may have *caught a break*(1), as **Samsung (NASDAQOTH: SSNLF)** has supposedly *pushed out*(2) its 3D NAND ramp after reports that it cut specialty equipment orders. **SanDisk**'s timeline doesn't include 3D NAND technology until the second half of 2015, whereas **Samsung** began mass production last August, and **Micron Technology (NASDAQ: MU)** intends to start sampling its 3D NAND products in the first half of the year.

Meanwhile, **SanDisk** *remains a leader*(3) in planar, or 2D, NAND solutions, providing the *lowest cost*(4). The company will ramp its 1Y (19nm x 19.5nm) technology this year, which made up just 15% of production in the fourth quarter of 2013, and transition to 1Z next year. Meanwhile, **Micron**'s 16nm technology has room for improvement as well as its 20.5nm in the bit line direction.

Both companies will likely squeeze as much as they can out of planar technology, and *stand to gain*(5) from any delays in **Samsung**'s ramp.

Why 3D? There's a physical limit to developing flash memory cells, and the *cost reduction from approaching that limit is not as great*(6) as earlier leaps in technology. As *returns on investment diminish*(7) for planar memory chips, chipmakers turned to stacking NAND strings vertically on top of each other. In this way, they can cram more transistors onto one die, thus *reducing the cost*(8) per bit.

But 3D *isn't more cost effective*(9) yet. **Samsung**'s first iteration that it started producing in August of last year featured 24 layers of stacked cells. The process uses just a 40nm, so there's room for improvement using a smaller node. But, the next generation of cell stacking is where *significant cost reduction*(10) will come in, as 24 layers grow to 32 layers or more.

3D is *more expensive to produce because it's more difficult*(11) to check a cell's operations when it's buried under a bunch of other cells. The process also requires additional specialty equipment to fabricate chips.

As a result, **SanDisk** believes it can get a *better return on investment*(12) through 2D NAND over an early transition to 3D NAND.

Such baseline can assist to evaluate the effect of a finer-grained analysis and determine the overall extent to which systems that incorporate information about the focus of the document perform better relative to a system that do not consider this information.

Using this algorithm, for any entity e amongst the set of entities E, and sentiment expression se among all the known SE, the entity sentiment relevance is set to be true. In the running example in Fig. 2, all sentiment expressions will be considered relevant to all mentioned entities.

This algorithm represents the standard mode of operation of document-level SA systems. In such systems, the final sentiment score for each entity would be the average sum of all sentiment expressions mentioned in the document. This method is expected to produce decent results when the document primarily discusses the entity of interest, or i.e., when focus of the document on the entity of interest is 'Main_topic'.

B. *Sentence-level analysis*

Sentence-level sentiment analysis systems consider every sentiment expression to be relevant if the corresponding sentence mentions the entity of interest or its co-references. So, for every sentence s in the document d, the entity sentiment relevance is set to be true if both e and se exist. In the example in Fig. 2, this algorithm will correctly catch all relevant sentiment expressions for the target entity SanDisk, but will also incorrectly set expression (2) as relevant. For the target entity Samsung, this algorithm will correctly catch sentiment expression (2), incorrectly catch expressions (1) & (5) and will not catch expressions (6)-(11). For the target entity Micron, this algorithm will correctly catch expression (5).

C. *Physical-Proximity-Based*

Physical-proximity-based algorithm is a text-range focused algorithm, which labels pieces of text as relevant or irrelevant according to how close they are to the references of the target entity and to other entities of the same type. This algorithm also uses some other contextual clues, such as, sentence and paragraph boundaries. In general, the mentioning of an entity starts its relevance range (and stops the relevance range of the previously mentioned entity). For the first entity reference in a paragraph, the range also extends backwards to the beginning of the sentence.

We looked into the following three variations to this algorithm, specifically adapted for different possible foci of the document on the entity of interest that were listed in section III:

'Minor_Topic_Proximity' – In this variation of the algorithm, all text is assumed to be irrelevant until the target entity is mentioned. This scenario is appropriate when analyzing an entity in documents not focusing on the particular entity. In Fig. 2, if the target entity is "Samsung", this algorithm will catch the sentiment expressions (2) & (10), since they follow a mention of the target entity reference in the same paragraph. It will also catch (9), since Samsung is the first entity mentioned in the paragraph and so its range extends backwards to the beginning of the paragraph. If the target entity is "SanDisk", the 'Minor_Topic_Proximity' algorithm will catch all 5 relevant expressions – (1),(3)-(5) & (12). Similarly, with "Micron" as the target, the algorithm will catch the one correct sentiment expression (5).

'Main_Topic_Proximity' – In this variation of the algorithm, all text is assumed to be relevant until another entity is mentioned. This scenario is suited for when the entity being analyzed is the main focus of the document. In Fig. 2, given that the target entity is "Samsung", this algorithm will catch all expressions caught by the above-mentioned 'Minor_Topic_Proximity' algorithm, plus expressions (6)-(8) & (11), since they appear in paragraphs without explicit non-target entity references. If the target-entity is "SanDisk", this algorithm will catch all 5 relevant expressions but will also catch four irrelevant sentiments: (6)-(8) & (11). With "Micron" as the target, this algorithm will catch the one correct sentiment, but will additionally catch four irrelevant sentiments, similarly to when "SanDisk" was the target entity.

'Few_Topics_Proximity' - In this variation of the algorithm, the relevance ranges interpolates over interchanging pieces of text, unless they are explicitly irrelevant (e.g., containing references of other entities of the same type). This scenario is suited when analyzing entities in documents equally focused on them. For the purposes of the Physical-Proximity algorithm, 'Related_Topics' is treated in the same way as 'Few_Topics'. Using the 'Few_Topics_Proximity' algorithm with "Samsung" as the target entity, the algorithm will catch (6)-(8), but not (11), because the next paragraph mentions the non-target entity "SanDisk", which interrupts interpolation. With "SanDisk" and "Micron" as the target entities, the algorithm 'Few_Topics_Proximity' will also catch all correct relevant expressions.

In our running example, altogether, 'Few_Topics' performs best on average for the three companies mentioned in this document. This is appropriate, since the three variations are specifically adapted for three types of entity foci within a document, and in the example, the entities are of focus 'Few_Topics'.

D. *Syntactic-proximity-based*

The second rule-based proximity algorithm we look into with regards to the focus of the document, is a syntactic-proximity-based algorithm. This algorithm is an expression-focused algorithm, which labels expressions as relevant or irrelevant according to their distance to various entity references in the dependency parse graph.

We look at two variations of the algorithm: 'direct' and 'reverse'. The former considers an expression relevant only if it is closest to the target entity from among all entities of the same type, and the distance is sufficiently close. The latter considers an expression irrelevant only if it has the above-described relation to some non-target entity of the same type. The rationale for the two variations is the distinction between 'Main_Topic' and 'Minor_Topic' document types with regards to the target entity. For the 'Minor_Topic' entities, a sentiment expression is assumed to be relevant only if it is explicitly connected to the entity, so the 'direct syntactic-proximity' is suitable. For 'Main_Topic' entities, an expression is irrelevant only if it is explicitly connected to another entity of the same type, so the 'indirect syntactic-proximity' is suitable.

In the running example in Fig. 2, all of the expressions (1), (3)-(5) & (12) are syntactically closely connected to references of the entity "SanDisk", so the direct variation would mark them as relevant to this entity. Expression (5) is directly connected also to "Micron". Consequently, the inverse variation would mark all expressions, except for the ones above, as relevant to "Samsung", which is correct in this case.

E. *Linear classification-based*

Instead of using proximity rules to determine whether a sentiment expression is relevant to a given entity, it is possible to use ML classification methods. We first consider a linear classification ML model, in which each candidate sentiment expression is an instance of a binary classification problem ("relevant" vs. irrelevant"), to be solved using supervised classification. For evaluating this algorithm, part of the test corpus is used for training, and the rest for testing, with N-fold cross-validation. The features for classification may use any information present in the input, such as references of target and non-target entities; appearances of paragraph and document boundaries; length of syntactic connections to target and non-target entities, when available; and explicit entity status within documents, when available.

In our experiments, the classifier takes as input the document text, labeled with all sentiment expressions, target entities, and non-target entities of the same type and returns a list of features for the classification instance, where each feature is a text string. So, the classifier:

- Receives, as input, a document text split into 'pieces', with a property list associated with each piece. For example, whether the piece contains sentiment expression, the target entity or non-target entity, paragraph break, the beginning or end of the document, number of links distance between the expression and target/non-target entity mention, etc.
- Outputs features that are sequences of properties of continuous pieces of length up to 6, one property from each piece, where one of the pieces must be the sentiment expression piece.

For the classification, we use a linear classifier with Large Margin training (regularized perceptron, as discussed in [31]).

For our experiments, which are described in the following section, we used the specific features and property lists that are outlined in Fig. 3. Of course, many more features can be thought to be incorporated, however, for the purpose of our experiments these appear to be sufficient.

Fig. 3. Feature and property lists used for the classification

- Sentiment expression → a single value 'SE',
- Mention of target entity → a single value 'TARGET'
- Mention of non-target entity → a single value 'OTHER'
- Paragraph break → a single value 'BR'
- Document start → a single value DOCSTART'
- Document end → a single value 'DOCEND'
- Text fragment → list of properties, among which there could be:
  - 'CAPS' if the fragment contains capital words;
  - 'NUMS' if the fragment contains numbers;
- Additional features 'SYNLENTARGET1', 'SYNLENTARGET2', 'SYNLENTARGET3', 'SYNLENOTHER1', 'SYNLENOTHER2', 'SYNLENOTHER3' track whether the sentiment expression is within distance of 1, 2, and 3 links from target and non-target entity mention.

F. *Sequence-classification-based*

The algorithm uses exactly the same features as the direct classification-based above, but instead of treating each relevance detection problem as a separate classification instance, they are placed in a sequence, according to how they appear within a document. Hence, instead of generating, for every document, many separate classification instance problems (one per each target entity and each sentiment expression), we generate several sequences (one sequence per target entity). In other words, we use a probabilistic sequence classifier (CRF, as discussed in [32] in place of a Large Margin binary classifier).

Engonopoulos et al. [4] also used CRF for entity level classification, but in their application, each segment is composed of a homogeneous sequence of sentiment labels that contains, at most, a single entity reference and the CRF classifier produces "sentiment flow" rather than a sequence of properties, as in our case.

G. Automatic identification of the focus of the document

As mentioned before, in addition to evaluating different algorithms in detecting sentiment relevance while using manually labeled information about the focus of the document, we evaluate them with such information that was labeled with supervised means. For this, we experiment with a classification-based algorithm for automatically identifying the focus of the document with regards to the different entities mentioned within it.

The classification problem instances are the pair 'Document' and 'Entity'. We used the same linear-classification model as described in section IV-E, but with a different set of features, as presented in Fig. 4 below.

Fig. 4. Feature used for automatically identifying the focus of the document

- ➢ Overall frequency of the entity within the document → IsThereOnlyOne, IsThereAtLeastThree, IsThereAtLeastFive, IsThereMoreThanThat
- ➢ Overall frequency of other entities of the same type within the doc.
- ➢ Whether the frequency of the entity is more or less than that of the other entities.
- ➢ Same as the first three features above, except that instead of counting entity instances, we count paragraphs in which the entities appear.
- ➢ Same as the first three features above, except that instead of counting the entity and the other entities of the same type, we count paragraphs in which the entity appears alone, and paragraphs in which it appears together with other entities of the same type.
- ➢ The maximal distance in paragraphs between two appearances of the entity, and again the comparison to other entities.
- ➢ Whether the entity appears in the document's title.
- ➢ All possible pairs of the features above.

Since our objective is primarily to demonstrate the feasibility of automatically identifying the focus of the document, we did not attempt to optimize the features and used features that sound applicable. Large Margin classifiers, such as the one we used in our experiments, are good at handling many irrelevant features, so the selected set is presumed to be adequate.

H. General settings

Each of the above-mentioned algorithms receives, as input, the text of the document, with labeled reference to the target entity and other entities of the same type. The labeled references also include all co-referential references, extracted automatically by an NLP (Natural Language Processing) system. The input text further includes labeled candidate sentiment expressions, which were either manually labeled or automatically extracted by a relevance-ignoring SA system. In our experiments, we also used a standalone automatic Financial SA system from [33], working in the 'ignore relevance' mode, which finds and labels all entities of the target type(s); resolves all co-references for the target entity type(s); finds and labels all sentiment expressions, regardless of their relevance; and provides dependency parses for all sentences in the corpus.

The task of the algorithms is to label each candidate expression as relevant or irrelevant to the target entity. The algorithms are evaluated in the next section according to the accuracy (precision, recall and F1) of this labeling of individual sentiment expressions. This method produces a reasonably well-understandable quality measure (the percentage of expressions that the algorithms get right or wrong), and also allows us to compare algorithms focused on individual expressions and algorithms working on text ranges.

V. EXPERIMENTS

In order to assess the value of information on the "focus of the document on the entity of interest", and to compare between the performances of different algorithms in recognizing sentiment relevance, with or without this information, we conducted several experiments. The initial results of these experiments were reported in [34]. In the current study, we first assessed the overall importance of relevance detection in SA by comparing the error levels in final sentiment polarities assignment in systems that do and do not utilize methods for sentiment relevance detection. Next, we established that information on the focus of the document on entities can be learned with supervised classification methods. We then evaluated the benefit of exploiting this information when determining sentiment relevance by selecting the appropriate variation of the algorithms accordingly. We conclude our experiments by determining the cross-domain applicability and evaluating the overall performance of the selected algorithms.

Fully annotating texts for sentiment relevance is an arduous task. In our experiments, we used two manually-annotated corpora[1] of a total of 320 documents, a financial corpus[2] of 160 financial news documents on at least one entity of interest and a medical corpus[3] of an additional 160 documents on a set of a few common drugs and diseases. Since this paper is primarily a study of sentiment relevance, the actual sentiment expressions are not always labeled in our datasets. Instead, relevance ranges are annotated for each entity, in the style of passage retrieval problems, with the expectation that sentiment expressions relevant to an entity only appear in the parts of the document that are labeled as "relevant", and conversely, that all expressions appearing in parts labeled "irrelevant" are irrelevant. This way of annotating allows the comparing of different relevance detection strategies, independent of the main sentiment extraction tool. All of the algorithms discussed in this paper use the same document processing methods, thus permitting us to compare the algorithms themselves independent of the quality and specifics of the underlying NLP. The evaluation metrics in all of the experiments are precision, recall, and F1. For the classification-based algorithms, unless stated otherwise, we use 10-fold cross-validation.

[1] Available at http://goo.gl/FTAugE.

[2] Average size ~5Kb. Mentions 424 different companies, in which the target entity type is COMPANY.

[3] Average size ~7Kb. Mentions 722 different people, 46 diseases, and 175 drugs, in which DISEASE, DRUG and PERSON are the target entity types.

A. Experiment: Importance of sentiment relevance

In the first experiment, we demonstrate the importance of identifying sentiment relevance when calculating the consolidated sentiment score for an entity within a set of documents. This experiment allows us to assess the overall impact of properly identifying sentiment relevance and can signal on the possible benefit of exploiting information on the focus of the document. We look at three particular scenarios, a baseline scenario which represents a document-level SA system that does not apply any sentiment relevance detection method, a main-topic scenario in which only documents that the focus on the entity of interest are deemed relevant for the SA, and a sentence-level SA scenario, which can be seen as a simple form of relevance detection, in which only sentences containing the entity of interest are considered relevant.

For each entity, we set the true sentiment score to the average of polarities of all relevant sentiments in a corpus. Then, we compare the true value with the values obtained using each of the above-mentioned scenarios. For the document-level scenario, we compared the true value with the average of polarities of all sentiments in all documents where the entity is mentioned. For the 'Main_Topic' scenario, we compared the true value to the average of polarities of all sentiments in the documents where the entity is labeled as the main-focused entity (the main topic of the document). This scenario models the typical state of a relevance-agnostic SA system. Finally, we compare the true value with the value obtained in the sentence-level scenario, which was calculated as the average of polarities of all sentiments in all sentences where the entity is mentioned.

For this evaluation, we only compare the sign of the final sentiment scores, without considering their magnitudes (unless it is close to zero, in which it is considered 'neutral'). The errors at this level indicate definite SA errors, i.e., miscalculating entity's sentiment into its opposite. The results of these experiments yielded that when using a document-level method, the error level in final sentiment polarity is 33% and 38%, for the financial and medial domains, respectively. When using a document-level method on a subset of 'Main_Topic' documents, the error level in final sentiment polarity drops to 12% and 28%, for the financial and medial domains, respectively. With a sentence-level method, the error-level is 17% and 22% for the financial and medical domains, respectively. These results indicate that information on the focus of the document is valuable for the overall task of SA. The 'Main_Topic' method naturally suffers from very low recall, with only 19% and 38% of entities covered in the financial and medical domains, respectively. As will be demonstrated in the next experiment, the low recall can be overcome by considering this method together with complementary methods that represent the other possible foci of documents. Similarly, when using a sentence-level system the results outperform a document-level system, this suggest that detecting sentiment relevant is, indeed, important.

B. Experiment: Automatic identification of document focus using classification

In the second experiment, we confirm that it is possible to identify the document focus with regards to it's different entities using supervised classification. For that, we used a subset of our annotated data for training, so the algorithm is trained-and-run-over the labeled set using 10-fold cross-validation, as described in section IV-G.The results of the direct evaluation, i.e., testing how well the model predicts the focus on the entity, given the training data, are as follows: when using 10-fold cross-validation, the results yield that the accuracy of supervised learning in the Medical and Financial corpora are 87.8% and 82.2% respectively. When using the Medical corpus for training and the Financial corpus for testing and vice versa, the accuracy is 78.2% and 86.1%, respectively. This validates that the supervised learning on cross-domain trainings data is possible. These results confirm that information extracted with supervised learning methods can be generalized to new language and new domains.

C. Experiment: Influence of document focus

After establishing that sentiment relevance detection has real value for the general task of SA and obtains a positive signal that the focus of the document could both be valuable information, which can also be learned with supervised classification means, we turn to evaluating how well various algorithms perform the task of sentiment relevance detection. In this set of experiments, we used both the manually annotated data and data, which was automatically labeled using supervised classification means. We compare between algorithms adapted to different possible documents foci when applied on subsets of documents of the corresponding types. We pay particular attention to two entity types: COMPANY in the financial corpus and DRUG in the medical corpus. We first evaluate the performances of the physical-proximity-based algorithm on the financial corpus by looking at how well each of the variations of this algorithm handle entities in different focused documents. For that, the set of all instances of relevance detection problems in the corpus (an instance consists of a sentiment expression within a text, together with a target entity) is divided into three subsets, according to the focus on the document with respect to the target entity. The results are shown in Table I.

TABLE I. PHYSICAL PROXIMITY ALGORITHMS' PERFORMANCES ON DIFFERENT SUBSETS OF DOCUMENTS WITH MANUALLY ANNOTATED FOCUS INFO (PRECISION/RECALL/F1) AND WITH SUPERVISED LEARNED FOCUS INFO (F1 (DIFF IN F1))

Algorithm	Focus information	'Minor_Topic' subset	'Main_Topic' subset	'Few_Topics' subset	Whole dataset
'Minor_Topic_Proximity'	Manually annotated	0.84/0.43/**0.57**	0.93/0.76/0.84	0.92/0.74/0.82	0.92/0.72/0.81
	Supervised learned	**0.60** (+0.026)	0.79 (-0.055)	0.83 (+0.011)	
'Main_Topic_Proximity'	Manually annotated	0.31/0.50/38	0.90/0.84/**0.87**	0.55/0.89/0.68	0.63/0.83/0.72
	Supervised learned	0.38 (-0.004)	**0.82** (-0.052)	0.73 (+0.043)	
'Few_Topics_Proximity'	Manually annotated	0.58/0.44/0.50	0.90/0.83/0.87	0.88/0.83/**0.86**	0.85/0.80/0.82
	Supervised learned	0.52 (+0.021)	0.81 (-0.059)	**0.87** (+0.016)	
'Combined_proximity'	Manually annotated				0.89/0.80/**0.84**
	Supervised learned				**0.83** (-0.012)

As can be expected, the three variations of the physical-proximity-based algorithm perform much better on the corpus subsets they are adapted to. Similarly, the combined-proximity algorithm that selects the appropriate variation of the proximity-based algorithm, shows an overall improvement in a corpus with mixed types of foci.

We can further learn from the table that the performances of the different variations of the algorithm on the appropriate subset of documents is significantly better both when dividing the dataset according to the manually annotated focus information and according to the supervised learn focus information. These results further illustrated the validity of supervised learned focus information.

Next, we evaluate the performance of the two variations of the Syntactic-proximity-based algorithm on the medical domain, having DRUG as the entity of interest. In this case, we divided the corpus to two subsets of 'minor' and 'major' topic focus, once when the focus was manually annotated and second when it was supervised learned. We used the 'direct' and 'inverse' variation of the syntax-proximity-based algorithm on both. The results are shown in Table II.

TABLE II. SYNTACTIC PROXIMITY-BASED ALGORITHMS' PERFORMANCES WITH MANUALLY ANNOTATED FOCUS INFO (PRECISION/RECALL/F1) AND WITH SUPERVISED LEARNED FOCUS INFO (F1 (DIFF IN F1)).

Syntactic-Proximity algorithm	**Focus information**	**'Main_topic'**	**'Minor_topic'**
Direct	Manually annotated	0.99/0.42/0.60	0.93/0.48/**0.64**
	Supervised learned	0.59 (-0.2%)	0.65 (+0.8%)
Inverse	Manually annotated	0.70/0.66/**0.68**	0.04/0.72/0.08
	Supervised learned	0.76 (+6.4%)	0.08 (-0.2%)

Same as with the proximity-based algorithm, the syntactic-proximity algorithm perform much better when its variation is adapted to the focus on the entity. These confirm the value of identifying the entity focus for syntactic-based algorithms.

We, additionally, compare in Table II the performance of the two classification-based algorithms on the two (whole) data sets, while either keeping or withholding the entity focus information from the classifier.

TABLE III. PERFORMANCES OF CLASSIFICATION-BASED ALGORITHMS (PRECISION/RECALL/F1).

Algorithm	**Focus info**	**Financial**	**Medical**
Linear Classification	None	0.931/0.875/0.902	0.852/0.844/0.848
	Manually annotated	0.919/0.904/0.912	0.910/0.844/0.871
	Supervised learned	0.920/0.903/0.911	0.903/0.860/0.881
Sequence Classification	None	0.967/0.853/0.906	0.934/0.857/0.894
	Manually annotated	0.975/0.875/0.922	0.968/0.876/0.920
	Supervised learned	0.972/0.869/0.918	0.949/0.882/0.914

As can be seen from the table, the difference in results is less pronounced here, but is still noticeable. The reason for the smaller difference, we hypothesize, is the ability of the classifiers to partially infer the entity focus from the various context clues that are used as classification features. In addition, the drop in performance when using supervised learned focus information is small, establishing the success of classification-based extraction of entity focus information.

D. *Experiment: Cross-domain applicability of classifications*

In this experiment, we test how well the classifiers trained on data from one domain work on input from a different domain. The classification results using different types of training data are shown in Table IV. Performance of classification-based algorithms using different training data (F1).

TABLE IV. PERFORMANCE OF CLASSIFICATION-BASED ALGORITHMS USING DIFFERENT TRAINING DATA (F1).

Training corpus	**Test corpus**	**Classification (10-fold)**	**Sequence Classification (2-fold)**
Medical	Medical	0.870	0.901
Financial,	Medical	0.869	0.902
Financial	Financial	0.917	0.905
Medical	Financial	0.911	0.907

The table confirms general independence of the classification performance on the domain. Comparing the 2-fold and 10-fold cross-validation results (in which the difference is equivalent to doubling the amount of training data), illustrates that the amount of training data is sufficient.

E. *Experiment: Overall performance of algorithms*

In this final experiment, we simply compare the overall accuracy of various algorithms for relevance detection, operating at their best parameters. The results are shown in Table V.

TABLE V. OVERALL PERFORMANCE OF DIFFERENT ALGORITHMS (F1).

Algorithm	**Financial**	**Medical**
Document-level (Baseline)	0.372	0.286
Sentence-level	0.800	0.774
Physical Proximity	0.841	0.795
Syntactic-Proximity	0.438	0.546
Classification	0.912	0.881
Sequence-Classification	**0.922**	**0.920**

As can be learned from the table, overall, classification-based algorithms perform better than the deterministic ones, with sequence-classification performing significantly better than direct classification. Syntactic-proximity-based is precise, but has relatively low recall, reducing its overall performance. The physical-proximity-based is the simplest, but produces reasonably high overall results, although worse than the best-performing classification-based methods.

VI. CONCLUSION

This paper explored the sentiment relevance detection problem and its solutions. We confirmed that relevance detection is essential for producing correct consolidated SA results and found that the focus of the document on the entity of interest is an important clue for solving the relevance detection problem. We also showed that this information can be effectively extracted automatically using supervised classification. By comparing several algorithms for relevance detection, we found that classification-based algorithms generally outperform simpler ones that are based on proximity, although even a very simple proximity-based algorithm performs reasonably well if provided with the entity focus information. The best single algorithm should, therefore, use all available proximity information, of all kinds, together with additional context information. Future work could include experimenting with different entity types from other domains and with additional ML methods for identifying the focus of an entity within a document. In addition, different feature sets can be used for the classification and the interactions between entities of the same type can further be studied with regard to sentiment relevance.

ACKNOWLEDGMENT

This research was supported by THE ISRAEL SCIENCE FOUNDATION (grant No. 906/13),by the Israel Ministry of Science and Technology Center of Knowledge in Machine Learning and Artificial Intelligence, by the Israel Ministry of Defense/ MAFAT and by the Intel Collaborative Research Institutes.

REFERENCES

[1] B. Liu, *Sentiment Analysis and Opinion Mining Synthesis Lectures on Human Language Technologies*. Morgan & Claypool Publishers., 2012.

[2] R. Feldman, "Techniques and applications for sentiment analysis," *Commun. ACM*, vol. 56, no. 4, p. 82, Apr. 2013.

[3] A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing HLT 05*, 2005, vol. 5, pp. 339–346.

[4] N. Engonopoulos, A. Lazaridou, G. Paliouras, and K. Chandrinos, "ELS: a word-level method for entity-level sentiment analysis," in *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, 2011, p. 12.

[5] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent twitter sentiment classification," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, 2011, pp. 151–160.

[6] N. Jakob and I. Gurevych, "Extracting opinion targets in a single-and cross-domain setting with conditional random fields," in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 2010, pp. 1035–1045.

[7] P. Turney, "Thumbs Up or Thumbs Down ? Semantic Orientation Applied to Unsupervised Classification of Reviews," in *Proceedings of the Association for Computational Linguistics (ACL)*, 2002, no. July, pp. 417–424.

[8] B. Pang and L. Lee, "A Sentimental Education : Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," 2004.

[9] E. Cambria, B. Schuller, B. Liu, H. Wang, and C. Havasi, "Knowledge-based approaches to concept-level sentiment analysis," *IEEE Intell. Syst.*, no. 2, pp. 12–14, 2013.

[10] E. Cambria, H. Wang, and B. White, "Guest editorial: Big social data analysis," *Knowledge-Based Syst.*, no. 69, pp. 1–2, 2014.

[11] E. Cambria, B. Schuller, B. Liu, H. Wang, and C. Havasi, "Statistical approaches to concept-level sentiment analysis," *IEEE Intell. Syst.*, no. 3, pp. 6–9, 2013.

[12] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up ? Sentiment Classification using Machine Learning Techniques," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Philadelphia, July 2002, pp. 79-86. Association for Computational Linguistics.*, 2002, no. July, pp. 79–86.

[13] A. Aue and M. Gamon, "Customizing sentiment classifiers to new domains: A case study," in *Proceedings of recent advances in natural language processing (RANLP)*, 2005, vol. 1, no. 3.1, pp. 1–2.

[14] H. Yu and V. Hatzivassiloglou, "Towards Answering Opinion Questions : Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2003, pp. 129–136.

[15] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 168–177.

[16] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Sentiment Analysis Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis," 2011.

[17] C. Scheible and H. Schütze, "Sentiment Relevance," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2013, pp. 954–963.

[18] B. Yang and C. Cardie, "Context-aware learning for sentence-level sentiment analysis with posterior regularization," in *Proceedings of ACL*, 2014.

[19] X. Ding, B. Liu, and P. Yu, "A holistic lexicon-based approach to opinion mining," in *The Conference on Web Search and Web Data Mining*, 2008, pp. 231–239.

[20] X. Liu and W. B. Croft, "Passage retrieval based on language models," in *Proceedings of the eleventh international conference on Information and knowledge management*, 2002, pp. 375–382.

[21] J. Tiedemann and J. Mur, "Simple is best: experiments with different document segmentation strategies for passage retrieval," in *Coling 2008: Proceedings of the 2nd workshop on Information Retrieval for Question Answering*, 2008, pp. 17–25.

[22] M. A. Hearst, "TextTiling: segmenting text into multi-paragraph subtopic passages," *Comput. Linguist.*, vol. 23, no. 1, pp. 33–64, Mar. 1997.

[23] J. Otterbacher, G. Erkan, and D. R. Radev, "Biased LexRank: Passage retrieval using random walks with question-based priors," *Inf. Process. Manag.*, vol. 45, no. 1, pp. 42–54, 2009.

[24] G. Salton, J. Allan, and C. Buckley, "Approaches to passage retrieval in full text information systems," in *Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval*, 1993, pp. 49–58.

[25] N. O'Hare, M. Davy, A. Bermingham, P. Ferguson, P. Sheridan, C. Gurrin, and A. F. Smeaton, "Topic-dependent sentiment analysis of financial blogs," *Proceeding 1st Int. CIKM Work. Top. Anal. mass Opin. TSA 09*, p. 9, 2009.

[26] D. Buscaldi, P. Rosso, J. Gómez-Soriano, and E. Sanchis, "Answering questions with an n-gram based passage retrieval engine," *J. Intell. Inf. Syst.*, vol. 34, no. 2, pp. 113–134, 2010.

[27] P. R. Comas, J. Turmo, and L. Màrquez, "Sibyl, a factoid question-answering system for spoken documents," *ACM Trans. Inf. Syst.*, vol. 30, no. 3, pp. 19:1–19:40, Sep. 2012.

[28] H. Wachsmuth, "Information Extraction as a Filtering Task Categories and Subject Descriptors," in *To appear in Proc. of the 22th ACM CIKM*, 2013.

[29] X. Ding and B. Liu, "The utility of linguistic rules in opinion mining," in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, 2007, pp. 811–812.

[30] M. Pontiki, H. Papageorgiou, D. Galanis, I. Androutsopoulos, J. Pavlopoulos, and S. Manandhar, "Semeval-2014 task 4: Aspect based sentiment analysis," *Proc. SemEval, Dublin, Irel.*, 2014.

[31] T. Zhang, "Regularized winnow methods," *Adv. Neural Inf. Process. Syst.*, pp. 703–712, 2001.

[32] J. Lafferty, A. McCallum, and F. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data.," in *Proceedings of the Eighteenth International Conference on Machine Learning (ICML-2001)*, 2001.

[33] R. Feldman, B. Rosenfeld, R. Bar-haim, and M. Fresko, "The Stock Sonar — Sentiment Analysis of Stocks Based on a Hybrid Approach," in *Proceedings of the Twenty-Third Innovative Applications of Artificial Intelligence Conference*, 2010, no. c, pp. 1642–1647.

[34] Z. Ben-Ami, R. Feldman, and B. Rosenfeld, "Entities' Sentiment Relevance," in *The 52nd Annual Meeting of the Association for Computational Linguistics*, 2014, vol. 2, pp. 87–92.