Enabling Advanced Business Intelligence in Divino

Danilo Croce, Francesco Garzoli, Marco Montesi,
Diego De Cao, and Roberto Basili

Department of Enterprise Engineering
University of Roma, Tor Vergata
00133 Roma, Italy
{croce,garzoli,montesi,decao,basili}@info.uniroma2.it

Abstract. This paper presents the system targeted in the Divino project, funded under the Industria 2015 framework of the Italian Ministry of Industry. The resulting platform embodies an innovative portal technology where Social Web functionalities, User Profiling and Aspect-based Opinion Mining are integrated through Liferay, a well known Enterprise Portal Technology. The proposed approach allows analysts to bootstrap an opinion-mining system by interacting with data-driven functions based on effective Online Machine Learning paradigms. The evaluation of the proposed methods is carried out in the targeted domain, i.e. the marketing of national wine products, one of the major focus area of the Made in Italy track of Industria 2015.

1 Introduction

In Business Intelligence, analysts have nowadays access to a variety of public forums where opinions and sentiments about companies, products and strategies are expressed in unstructured form. Opinion Mining (OM) [11] tackles different problems that arise in this scenario, such as determining if a segment of text (sentence, paragraph or section) is *opinionated*, identifying the *opinion-holder* (the person or organization who expresses the opinion) or determining the *polarity* (i.e. how positive or negative each opinion is). For business intelligence, it is also useful to classify each opinion according to the *aspect* of the analyzed product, such the flavor or taste of a wine.

This paper discusses the system targeted in the Divino project, funded under the Industria 2015 framework of the Italian Ministry of Industry. The resulting platform embodies an innovative portal technology where Social Web functionalities, User profiling and Aspect-based Opinion Mining (OM) are integrated. On the one hand, users can visit a portal enjoying a community interested in the eno-gastronomic domain of wine. When logged, the so-called Divino User has a deeper interaction with the portal, leaving message in the forum, designing a personalized blog or buying items in a specialized e-shop; every registered user becomes part of a Social Network, determining friendship-based links with

other users. On the other hand, an Opinion Mining workflow has been implemented to capture people opinions and preferences expressed within the portal. These are enriched by crawling and processing specialized sites and blogs from the Web. Opinions are stored in a semi-structured form and meaningfully summarized to be consumed by Market Analysts. Based on the Enterprise Portal Technology known as Liferay, the system results in a Web Portal where different users can enjoy and interact, always providing valuable information for Business Intelligence processes.

The proposed OM workflow is quite general and it can be used to bootstrap and adapt an OM system to a target domain. This can be achieved by applying online Learning Algorithms [3], training classifiers that recognize topics, aspects and opinions in texts, comments and blogs. The online learning paradigm is appealing as it allows an interaction between the system and a Market Analyst, who can incrementally refine the domain by validating classifiers predictions. The applicability of the proposed approach is then evaluated in the targeted domain of the national and international marketing of wine products, one of the major focus area of the Made in Italy track of Industria 2015. In the rest of the paper, Section 2 discusses the OM process in Divino. Section 3 provides a description of the resulting portal. Section 4 provides the experimental evaluation and Section 5 derives the conclusions.

2 Modeling Opinion Mining in Divino

If we are interested in detecting opinions about wines, all textual units containing information related to the target products must be carefully retrieved. Let us consider the following excerpt related to the wine domain:

> *La gamma aziendale prevede un vino rosso basato su uve ciliegiolo in purezza, il Ciliegiolo Golfo del Tigullio doc, vinificato in acciaio, che dona al vino netti ma delicati sentori di ciliegia, violetta e una sottile vena speziata (pepe) senza mancare di una buona aciditá e tannicitá.[1]*

It contains information about a wine, the "*Ciliegiolo Golfo del Tugullio doc*", i.e. the entity to which the author refers. As we are interested in opinions related to specific aspects of wine, such as *flavor* and *taste*, textual units containing objective expressions can be neglected. Words like "*sentori netti ma delicati*" and "*buona aciditá e tannicitá*" here give a positive connotation to the *Aroma* and *Taste* aspects, respectively. Moreover, even if not made explicit, the underlying domain must be properly addressed as it allows to reject texts related to other products, e.g. cars or mobile phones.

Many approaches have been defined to determine and recognize opinions in texts, as discussed in [8, 11, 14], ranging from different text genres, from newswire

---

[1] Translation: *The product range contains a red wine derived from Ciliegiolo grapes, that is the Ciliegiolo Golfo del Tigullio doc, vinified in stainless steel, which gives strong but delicate hints of cherry, violet and a slightly spicy note (pepper) without missing a good acidity and tannin levels.*

[17] to social media, such as Twitter [10]. These studies let to the development of several corpora with detailed opinion and sentiment annotations, e.g., the MPQA corpus [16] of newswire text. These corpora have proved very valuable as resources for learning about the language of sentiment in general. As discussed in the following section, in Divino we applied empirical methods in order to automatically train classifiers able to associate sentences to specific classes useful to characterize the writer opinion. More formally, our ultimate aim is therefore to extrapolate structured information such as the n-tuple $\langle u, t, h, r, a, b \rangle$ where:

- u is the **Textual Unit**, e.g. a sentence or paragraph expressing an opinion;
- t is the **Topic** related to u, e.g. the `WineryProduct`, that represents the opinion domain;
- h is the **Opinion Holder**, the person or organization expressing the opinion (here the blog author);
- r is the **Opinion Target**, that is the entity subjectively valued (e.g. *Cilie-giolo Golfo del Tugullio doc.*);
- a is the **Aspect** for r in the domain t (e.g. *flavor* or *taste*);
- b is the **Polarity**, associated with a target r and its specific aspect a, e.g. Positive, Negative or Neutral.

In the next section, data-driven learning algorithms to associate each u to the proper n-tuple will be discussed.

2.1 The Opinion Mining Workflow

Behind the Divino portal, an OM workflow has been developed to structure opinions, as discussed above. We defined a specific ontology providing a meta-model from which domain-specific OM workflows are derived, not shown here for space reasons. In the Divino project, the workflow shown in Figure 1 has been implemented.

In the **Data Gathering** phase, a dedicated *Web Crawler* downloads documents from wine specialized sites, blogs and forums. *Chaos* [1], the *Natural Language Processing* (*NLP*) processor made available at the University of Tor Vergata, analyzes such documents to extract morpho-syntactic and semantic information required by the workflow.

In the **Information Extraction** phase the *Target Extractor* allows to identify sentences mentioning one or more target products. In the domain addressed by Divino, examples of target can be wines, such as *Barolo* or *Taurasi*, or Varietal, such as *Syrah* or *Merlot*. This module is based on the Name Entity Recognizer and Classifier (NERC) made available by Chaos. The *Target Propagator* finds sentences referring to targets, even if they are not explicitly mentioned.

The core **Sentiment Analysis** functionalities determine opinions and are realized as a sequence of classification steps. Among all existing Machine Learning paradigms, we investigated the class of Online Learning Algorithms. The goal, as in a traditional fashion, is to predict classes for instances. In addition, soon after the prediction is made, it can then be used to refine the prediction hypothesis used by the algorithm. In a traditional setting, the training phase would
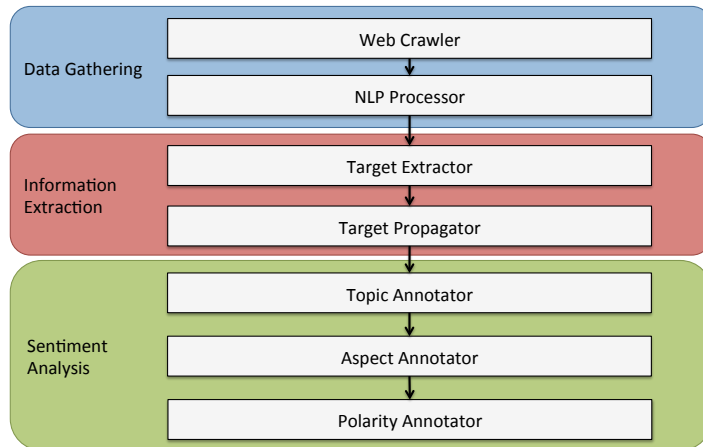
Fig. 1: The OM workflow

have started *ex-novo*, re-considering all training examples. Such online schemas allow implementing mechanisms for relevance feedback: it incrementally refines the domain classifiers and adapts the resulting analysis to the target domain.

In particular, the Passive Aggressive (PA) learning algorithm [3] is one of the most popular online approaches and it is generally referred as a state-of-art online method. Its core idea is quite simple: when an example is misclassified, the algorithm updates the model with the hypothesis that is more similar to the current one. Formally, let $(\mathbf{x}_t, y_t)$ be the t-th example where $\mathbf{x}_t \in \mathbb{R}^d$ is a feature vector that represents a document or sentence in a d-dimensional space, while $y_t \in \{+1, -1\}$ is the corresponding label, e.g. a sentence does/does not belong to a topic or polarity class. Let $\mathbf{w}_t \in \mathbb{R}^d$ be the current classification hypothesis. The PA classification function is $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$. After receiving $\mathbf{x}_t$, the new classification function $\mathbf{w}_{t+1}$ becomes the one that minimizes the objective function $Q(\mathbf{w}) = \frac{1}{2} \|\mathbf{w} - \mathbf{w}_t\|^2 + C \cdot l(\mathbf{w}; (\mathbf{x}_t, y_t))$. The first term $\|\mathbf{w} - \mathbf{w}_t\|$ is a measure of how much the new hypothesis differs from the old one while the second term $l(\mathbf{w}, (\mathbf{x}_t, y_t))$ is a proper loss function assigning a penalty cost to an incorrect classification. C is the aggressiveness parameter that balances the two competing terms[2]. Minimizing $Q(\mathbf{w})$ corresponds to solving a constrained optimization problem, whose solution let to update the classifier according to the following schema: $\mathbf{w}_{t+1} = \mathbf{w}_t + \alpha_t\mathbf{x}_t, \ \ \alpha_t = y_t \cdot \min\left\{C, \frac{H(\mathbf{w}_t; (\mathbf{x}_t, y_t))}{\|\mathbf{x}_t\|^2}\right\}$.

If $\mathbf{x}_t$ is correctly classified, the model does not change, while, after a wrong prediction, the new classification function $\mathbf{w}_{t+1}$ becomes a linear combination between the old one $\mathbf{w}_t$ and the feature vector $\mathbf{x}_t$. A kernelized version of the PA algorithm is easy to obtain and gives the possibility to exploit rich data representations, as discussed in [3, 15].

---

[2] In this work we will consider the hinge loss $H(\mathbf{w}; (\mathbf{x}_t, y_t)) = max(0, 1 - y_t\mathbf{w}^T\mathbf{x}_t)$

Fig. 2: Forum and news portlet for non registered users.

In the resulting workflow, given a new document, the *Topic Annotator* retrieves paragraphs related to all topics t that are compatible with the domain, e.g. `WineryProducts` or `Varietals`. Each paragraph is associated by a PA classifier to each target topic t. In order to model an open-world scenario, where not all topics are already known, the `OtherTopic` class is introduced: each paragraph classified as `OtherTopic` is not considered in the remaining processing chain by the other annotators. The *Aspect Annotator* classifies all sentences from the remaining paragraphs with respect to the active aspects a of a given topic t. Even at this level, the open-world assumption is valid, so the `OtherAspect` class is introduced. Finally, for each sentence associated to a valid aspect, the corresponding polarity is provided by another PA-based classifier with respect to the POSITIVE, NEGATIVE or NOPOLARITY classes[3]. More details about the modeling of single textual units u are provided in Section 4. At the moment of writing the Opinion Holder h is assumed to be the content creator, e.g. the author of a blog page or comment in a forum.

3 The Divino portal

The Divino portal is designed as a set of interacting services whose overall logic is integrated within the Liferay portal. Liferay[4] is a free and open source enterprise portal written in Java and distributed under the GNU Lesser General Public License and proprietary licenses. It allows to efficiently create a portal for Internet or Intranet use and it is fundamentally constructed of functional

---

[3] When a sentence is classified as POSITIVE and NEGATIVE at the same time, it is considered as NEUTRAL.
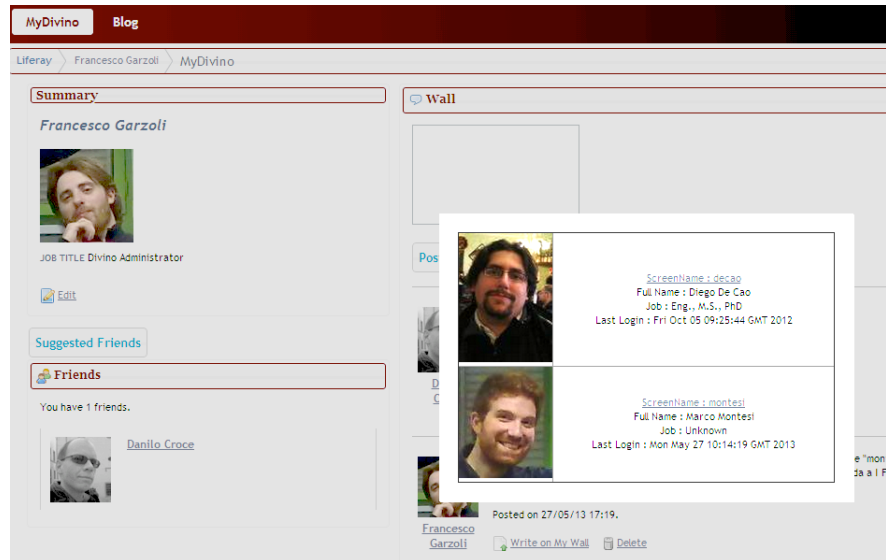
[4] http://www.liferay.com/

Fig. 3: The MyDivino page: in evidence the Friend Suggestion

units called *portlets*, that represent portal functionalities and produce fragments of markup code that are aggregated into a portal.

Liferay enables the creation of different users and different roles, so that every role associates a user group sharing the same permissions. Permissions are linked to Portal, Portlet and other Liferay entities. In addition to the role of Administrator, the Divino Portal handles four roles, i.e. Guest, Divino User, Annotator and Market Analyst, each enabled to access to the following functionalities.

Enjoying the Divino Portal as a Registered User. A user can visit the Divino Portal without being registered. As a *Guest*, he can view a limited set of pages providing not tailored information as well as the *Divino Forum* and e-commerce services, i.e. the *Divino Shop*, as shown in Fig. 2. A log-in step is required in order to post any message or buy items. Moreover, a *Divino Search* portlet allows to retrieve all web pages downloaded during the Data Gathering phase, described in Section 2. When logged-in at the Divino Portal, the user assumes the role of *Divino User*. He can now participate to the social activities made available in the portal within the forum and e-commerce portlets. As shown in the background of Figure 3, each user is associated to a personal *MyDivino* page where a blog can be easily populated with comments. In line with popular Social Networks, a friendship schema is applied to allow a restricted number of friends to read the personal blog. Each user can retrieve other users and ask their friendship. Every Divino User owns a profile that keep all the information about his search queries, preferences and purchased items. Such interactions with the system, as well as other information provided through a questionnaire suggested in the registration phase, are crucial for many portal functionalities. They enable

Fig. 4: Annotation interface

the design of different User Recommending and Information Filtering schemas, as discussed in [12]. At the moment of writing, a first recommending schema is used to suggest friends. All information gathered during registration provide a set of preferences P_{u_i} describing each Divino User d_i. For example, one can prefer red wines instead of white wines or wines from specific regions. A first recommending function has been implemented by estimating the similarity among user pairs d_i and d_j in terms of the Jaccard Similarity score between the sets of related preferences: $J(d_i, d_j) = \frac{|P_{d_i} \cap P_{d_j}|}{|P_{d_i} \cup P_{d_j}|}$. The score is 1 for user pairs with exactly the same interests, while it drops to 0 for "different" users. Figure 3 shows the User Suggestion, i.e. two users nominated to be friends.

Providing labeled material as Divino Annotator. The machine learning methods proposed in Section 2.1 require labeled data in order to acquire a proper model of target phenomena. The role of *Divino Annotator* allows user to access the annotation functionalities. When logged, users can retrieve, add, remove and modify documents downloaded during the Data Gathering phase. Given a document, the user annotates all paragraphs with the corresponding information, such as Topic, Aspects and Polarity. In Figure 4 the interface shows a brief part of a document related to a specific wine, the *Chianti Classico*: in particular, two sentences expressing positive comments about the *taste* aspect are shown. The contribution of the Online Learning schema is emphasized in the annotation phase. In fact, the annotator can ask the system to automatically annotate the examples and validate the proposed information. When these are validated and submitted, the model can be corrected and improved through the novel annotations, so conforming to the Annotator notion of the target domain. In a real scenario, the system is expected to produce wrong annotations during its
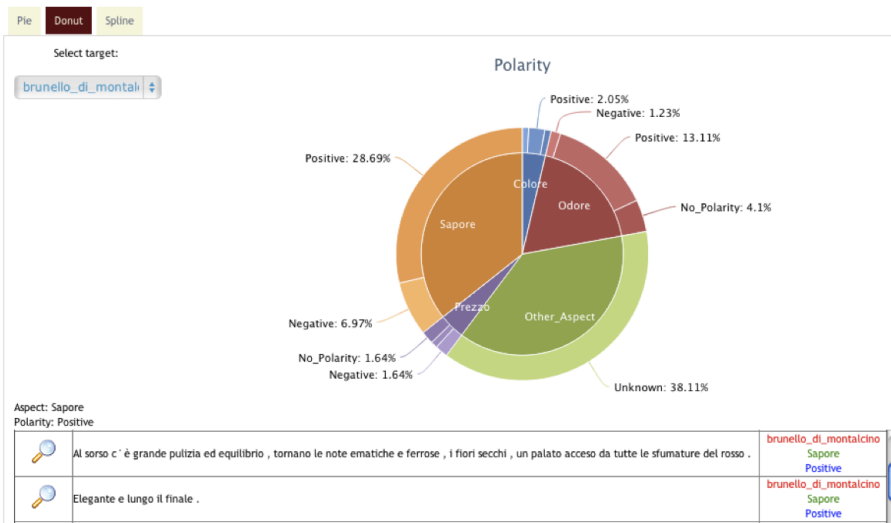
Fig. 5: Market Analyst interface

first life-cycle and to improve the annotation quality after a reasonable number of interactions with the annotators.

Advanced Business Intelligence in Divino. The automatic detection of users preferences and opinion from the portal and the corresponding translation in a semi-structured form, represents a valuable source of information for *Market Analysts* to feed Business Intelligence processes. Some of these information are automatically captured from user interactions, while other come from external sources, retrieved in the Data Gathering phase. For example, the Market Analyst can browse statistics about purchased items or the query logs from the Search Portlet. Advanced Business Intelligence techniques can also be applied in order to capitalize the knowledge extracted within the Opinion Mining process, as discussed in [11]. As an example, the Opinion Browsing portlet is shown in Figure 5: a multi-level pie chart, the so-called Donut, provides a synthetic view of opinions expressed by people within the forum or the targeted web pages. It is represented as the percentage of textual units expressing opinions about different aspects within a specified domain, such as `WineryProduct`. A fine-grained analysis can be enabled focusing on a specific target, e.g. a *Brunello di Montalcino*. For example, in Figure 5 the percentage of textual units giving positive comments about the *taste* of the product is 29,69%, while the percentage of negative comments about the *price* is 1.64%. The analyst can have a deep look on these statistics by clicking on every percentage, so visualizing the list of textual units and, if needed, can browse the source document. It is also possible to access to other reports and charts, enabling complex activities such as the monitoring of temporal trends, by visualizing the opinion depending on specific temporal based selections.

4 Experimental Evaluation

In this section, the Opinion Mining process is evaluated, as it represents the core functionality enabling Advanced Business Intelligence processes within the entire Divino Portal. In particular, the quality of classifiers powering different annotators described in section 2.1 is considered. The classification task is tackled through a Multiple Kernel approach, as discussed in [15]. Kernel methods are beneficial because the combination of kernel functions can be integrated into state-of-the-art classifiers, such as Support Vector Machines [15] or Passive Aggressive algorithm [3], as they are still kernels.

4.1 Textual Unit representation

A multiple kernel approach allows to combine the contribution of complex kernel functions to implicitly integrate different linguistic and semantic information of annotated examples. In this work, two kernels have been employed in our modeling. The Bag of Word Kernel (BOWK) reflects the lexical overlap between textual units t, represented as a vector whose dimensions correspond to different words. Each dimension represents a boolean indicator of the presence or not of a word in the text. The kernel function is the cosine similarity between vectors.

Another kernel is added, as lexical information of BOWK is highly affected by data sparseness, and words as found in test cases may often result rare or unseen in the training set. Our aim is to increase robustness to the resulting system by extending lexical information through *Distributional Analysis*. The core idea is that the meaning of a word can be described by the set of textual contexts in which it appears (*Distributional Hypothesis* as described in [6]). Words can be geometrically represented as vectors whose components reflect the corresponding contexts: two words close in the space (i.e. they have similar contexts) are likely to be related by some type of generic semantic relation, either paradigmatic (e.g. synonymy, hyperonymy, antonymy) or syntagmatic (e.g. meronymy, conceptual and phrasal association), as observed in [13]. A word-by-context matrix M is obtained through a large scale corpus analysis. Then the *Latent Semantic Analysis* [9] technique is applied to capture the statistical information of M by a lower k-dimensional space. Given two words w_1 and w_2, their similarity function σ is estimated as the cosine similarity between the corresponding projections $\boldsymbol{w_1}, \boldsymbol{w_2}$ in the space, i.e $\sigma(w_1, w_2) = \frac{\boldsymbol{w_1} \cdot \boldsymbol{w_2}}{\|\boldsymbol{w_1}\|\|\boldsymbol{w_2}\|}$. The result is that every word can projected in the reduced Word Space. The representation of a t consists of a linear combination of vectors representing words. Finally, the resulting Lexical Semantic Kernel (LSK) function is the cosine similarity between vector pairs, in line with [4], allowing to generalize the lexical information. The Word Space is acquired through the distributional analysis of a corpus made of about 2.5 million tokens; it is composed by web pages downloaded during the Data Gathering phase and pages from Wikipedia related to the `Wine` category, in order to have a space tied to the target domain. All words occurring more than 30 times are represented through vectors. The original space dimensions are generated from the set of the 20,000 most frequent words in the corpus. One dimension describes

Fig. 6: Examples of tag cloud: words referring positively to the *taste* of wine and negatively to the *price* of wine are shown on the left and right, correspondingly.

the Pointwise Mutual Information score between one feature, as it occurs on a left or right window of 5 tokens around a target. Left contexts of targets are treated differently from the right ones, in order to capture asymmetric syntactic behaviors (e.g., useful for verbs): 40,000 dimensional vectors are thus derived for each target, later reduced to $k = 250$.

As a side effect of the LSK, sentences are projected in the same representation space of words as in [9]. Given a textual unit u referring to an aspect a with a polarity p, the set of m words more semantically related to us can be obtained, namely W_t^k. By collecting all $W_{a_p}^k$ from sentences referring to a specific aspect a with a polarity p, a Tag Cloud can be obtained, as discussed in [5]. Figure 6 shows tag clouds related to the *taste* and *price* aspects. They are generated by selecting the $k = 20$ words more similar to examples used in this experimental evaluation. Notice that the word size depends on the number of times a tag is suggested by a single u.

4.2 Opinion Mining Results

In our approach, the kernel combination αBOWK $+$ βLSK estimates the similarity between textual units, linearly combining lexical properties captured by BOWK and the lexical generalization of the LSK[5]. A set of 60 web pages has been annotated according to the schema proposed in Section 2.1. Annotations are derived from 7 specialized sites and blogs[6] from the enogastronomic domain targeted in the Divino Project. The Topic annotator is powered with a classifier associating paragraphs with respect to 4 classes, i.e. WINERYPRODUCTS, VARIETALS, WINERYBRANDS and OTHERTOPICS. The analysis has been then specialized for the WINERYPRODUCTS and each sentence within this topic has been classified with respect to different aspects, i.e. TASTE, AROMA, COLOR, PRICE and OTHERASPECTS. Each sentence related to a valid aspect is then

---

[5] Here, parameters α and β weight the combination of the three kernels. In our experiments, α and β are set to 1.

[6] We annotated pages from `www.intravino.com`, `www.enofaber.com`, `percorsidivino.blogspot.it`, `ilvinoeoltre.blogspot.it`, `grappolidivini.blogspot.it`, `simodivino.blogspot.it` and `grappolorosso.blogspot.it`.

Table 1: Number of examples for the Topic, Aspect and Polarity classifiers.

Topic	#par.		Aspect	#sent.		Polarity	#sent.
WINERYBRANDS	71		AROMA	222		POSITIVE	411
WINERYPRODUCTS	293		COLOR	60		NEGATIVE	77
VARIETALS	25		PRICE	24		NOPOLARITY	141
OTHERTOPICS	42		TASTE	323		Total	629
Total	431		OTHERASPECTS	239			
			Total	868			

classified with respect to the POSITIVE, NEGATIVE and NOPOLARITY classes. Table 1 shows the number of paragraphs annotated with Topic classes and the number of sentences annotated with Aspect and Polarity classes.

Table 2: Accuracy of the SVM and PA classifiers in the 10-fold cross validation schema; in parenthesis the standard deviation is reported.

	SVM	PA
Topic	85.45% (5.14%)	83,53% (7,44%)
Aspect	85.25% (3.60%)	84,67% (7,02%)
Polarity	79.17% (5.63%)	75,18% (15,89%)

In order to evaluate the robustness of the employed Passive Aggressive (PA) classifiers, we compared performances with a Support Vector Machine based classifier, which represents the state-of-the-art of kernel-based (non online) machines. In particular, the $SVM^{multiclass}$ schema described in [7] is applied[7]. A One-VS-All schema is used for the PA to realize the multi-classification: a binary classifier is used for each class and the one providing the highest classification function is selected. As the PA model depends on the order of example provided in the training phase, a 10 fold cross validation schema is applied. On the contrary, $SVM^{multiclass}$ adopts the implicit multi-class formulation described in [2]. Results are measured in terms of accuracy, i.e. the percentage of examples obtaining the correct labeling. Table 2 shows the mean results of both classifiers within the 10 folds. As expected, the SVM generally achieves slightly higher and more stable scores. It is not surprising as SVM, as a batch learning algorithm, finds the optimal solution of the classification problem, while the PA does not, according to its online nature [3]. However, high results achieved by different PA classifiers, i.e. about the 80% accuracy, confirms the applicability of online schema in the OM workflow within the Divino Portal. The slightly lower accuracy of the polarity classifiers emphasizes the complexity of capturing opinions in the domain of wine.

---

[7] http://svmlight.joachims.org/svm_multiclass.html

5   Conclusion

This paper shows a comprehensive web portal where Social Web functionalities, User Profiling and automatic Aspect-based Opinion Mining are integrated. The resulting portal allows people to express their preferences while enabling Market Analysts to bootstrap an opinion-mining system from scratch. The effectiveness of the proposed Online Machine Learning schema has been evaluated in a real use case in the national marketing of wine products. Future work will focus on improving the system bootstrapping capability with fewer annotated data, as well as a deeper study to combine modern Business Intelligence to semi-structured information extracted through Opinion Mining techniques.

References

1. Basili, R., Zanzotto, F.M.: Parsing engineering and empirical robustness. Nat. Lang. Eng. 8(3), 97–120 (Jun 2002)
2. Crammer, K., Singer, Y.: On the algorithmic implementation of multi-class svms. Journal of Machine Learning Research 2, 265–292 (2001)
3. Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., Singer, Y.: Online passive-aggressive algorithms. Journal of Machine Learning Research 7, 551–585 (2006)
4. Cristianini, N., Shawe-Taylor, J., Lodhi, H.: Latent semantic kernels. J. Intell. Inf. Syst. 18(2-3), 127–152 (2002)
5. Halvey, M.J., Keane, M.T.: An assessment of tag presentation techniques. In: Proceedings of WWW 2007. pp. 1313–1314. ACM, New York, NY, USA (2007)
6. Harris, Z.: Distributional structure. In: Katz, J.J., Fodor, J.A. (eds.) The Philosophy of Linguistics. Oxford University Press (1964)
7. Joachims, T., Finley, T., Yu, C.N.: Cutting-plane training of structural SVMs. Machine Learning 77(1), 27–59 (2009)
8. Kim, S.M., Hovy, E.: Determining the sentiment of opinions. In: Proceedings of COLING. Association for Computational Linguistics (2004)
9. Landauer, T., Dumais, S.: A solution to plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. Psychological Review 104 (1997)
10. Nakov, P., Rosenthal, S., Kozareva, Z., Stoyanov, V., Ritter, A., Wilson, T.: Semeval-2013 task 2: Sentiment analysis in twitter. In: In SemEval 2013. pp. 312–320. Atlanta, Georgia, USA (June 2013)
11. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2(1-2), 1–135 (Jan 2008)
12. Rajaraman, A., Ullman, J.D.: Recommendation Systems, chap. 9. Cambridge University Press (2011)
13. Sahlgren, M.: The Word-Space Model. Ph.D. thesis, Stockholm University (2006)
14. Seerat, B., and, F.A.: Article: Opinion mining: Issues and challenges (a survey). International Journal of Computer Applications 49(9), 42–51 (July 2012), published by Foundation of Computer Science, New York, USA
15. Shawe-Taylor, J., Cristianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press, New York, NY, USA (2004)
16. Wiebe, J., Wilson, T., Cardie, C.: Annotating expressions of opinions and emotions in language. Language Resources and Evaluation 1(2), 0 (2005)
17. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of EMNLP. Stroudsburg, PA, USA (2005)