Looking for Opinion in Land-Use Planning Corpora

Eric Kergosien, Cédric Lopez, Mathieu Roche, Maguelonne Teisseire

Looking for Opinion in Land-use Planning Corpora

Eric Kergosien[1,2], Cédric Lopez[4],
Mathieu Roche[3], and Maguelonne Teisseire[1,2]

[1] LIRMM - CNRS, Univ. Montpellier 2,
161 rue Ada, 34095 Montpellier Cedex 5, France
eric.kergosien@lirmm.fr
[2] Irstea, UMR TETIS, 500 rue J.F. Breton, 34093 Montpellier Cedex 5, France
maguelonne.teisseire@teledetection.fr
[3] Cirad, UMR TETIS, 500 rue J.F. Breton, 34093 Montpellier Cedex 5, France
mathieu.roche@cirad.fr
[4] Viseo - Objet Direct, 4 av. du Doyen Louis Weil, 38000 Grenoble, France
clopez@objetdirect.com

Abstract. A great deal of research on opinion mining and sentiment analysis has been done in specific contexts such as movie reviews, commercial evaluations, campaign speeches, etc. In this paper, we raise the issue of how appropriate these methods are for documents related to land-use planning. After highlighting limitations of existing proposals and discussing issues related to textual data, we present the method called Opiland (OPinion mIning from LAND-use planning documents) designed to semi-automatically mine opinions in specialized contexts. Experiments are conducted on a land-use planning dataset, and on three datasets related to others areas highlighting the relevance of our proposal.

Keywords: Land-use planning, Text-Mining, Opinion-mining, Corpus, Lexicon.

1 Introduction

The notion of territory, and more specifically of land-use-planning, is complex and refers to many concepts such as stakeholders, spatial and temporal features, opinions, politics, history, etc. Hence, territories reflect both economic, ideological and political appropriation of space by groups who provide a particular view of themselves, of their history, and their uniqueness. The characterization and understanding of perceptions of a territory by different users is complex but needed for land-use planning and territorial public policy. In this paper, we focus on on political and administrative territories (eg. local or regional territory) and we propose an original approach to build specific vocabularies of opinions related to our domain.

Opinion mining has been intensively studied in various fields such as movie reviews, political articles, tweets, etc. Methods are based on statistics or Natural Language Processing (NLP). A lexicon or a dictionary of opinions (with or without polarity) is often used. In the context of land-use planning, even if information published on the web (blogs, forums, etc.) and in media expresses a feeling, the traditional opinion mining approaches fail to extract opinion due to the context specificity (small or medium-size and specialized corpus). We propose to tackle this issue by defining a new approach, called Opiland (OPInion mining for LAND-use planning documents), in order to semi-automatically mine opinion in specific contexts. This approach uses specialized vocabularies to compute a polarity score for documents.

This paper is structured as follows. In Section 2, an overview of opinion mining methods is presented. In Section 3, the Opiland method is detailed. Section 4 reports experiments, firstly on the land-use planning corpus and secondly on three other corpus. The paper ends with our conclusions and future work.

2 State-of-the-art

In **opinion mining** and NLP, the analysis of subjectivity and opinions expressed by people in texts (newspapers, technical documents, blogs, reviews, letters, etc.) is called opinion analysis [12]. Recognition polarity attempts to classify texts according to positivity or negativity with respects to the expressed opinions therein. Two main approaches can be identified: one based on the frequency of positive and negative words in each text [19], and the other one based on machine learning techniques from annotated texts [6]. Hybrid approaches would appear to offer the best results [9, 10]. In all these approaches, several features are used, including words, n-grams of words [13], the shifted words [8], and so on. These features can be exploited using machine learning methods based on an annotated corpus. Such corpora are made available in text analysis challenges such as TREC (Text Retrieval Conference), or DEFT (DÉfi Fouille de Textes) for assesment by the French community. However, only a few are annotated according to opinion and polarity. In addition, several classification methods can be grouped into voting systems proposed by [17] or applying reinforcement and bag of words methods [7]. Other approaches rely on incremental methods for opinion analysis [21]. Along with the classification of opinion texts, the team's research works focused on the automatic construction of opinion vocabularies [5]. The incremental approaches proposed are usually based on web-mining methods in order to learn an opinion vocabulary specifically linked to a topic or a subtopic.

Concerning **analysis of user feelings** in the **land-use planning** domain, the overview of existing research reveals the involvement of several communities. For thirty years, the concept of territory, based on different definitions, has been widely used and discussed by ethologists and ecologists, geographers, sociologists, economists, philosophers, etc. In the French community, the geographer community has been particularly prolific, and their work has tended to adopt either a social or a political angle in their territorial analysis [20]. On the one hand, social geography analyzed the identity dimension of the territory, the membership reports and tighten reports [2]. On the other hand, political geographers have tried to represent dimensional aspects of the territory, through the analysis of public action initiatives [4]. To cope with its multiple definitions of territory, it often clarifies the words meaning by adding a qualifier: Biophysical territories (watershed, great landscape, etc.), politico-administrative territories (e.g. city, country, continent, etc.), large territories, suitable territories, mobile territories. According to the state-of-the-art, different types of approaches could be used. In our context, supervised approaches are not adapted because we process with a reduced size of labeled data (i.e data analysed by experts). Unsupervised approaches based on incremental methods using seed of polarized words to enrich are often used in sentiment analysis studies [16]. But they are not adapted in our context. Generally the enrichment is based on global information (e.g. hits of Web pages) returned by general search engines (for example, Google, Yahoo, Exalead, and so forth). But this is too general in our context.

However, to the best of our knowledge, there is no automatic or semi-automatic method to mine opinion in the land-use planning context. In this paper, the proposed approach defines a specific vocabulary of opinion related to the domain using general lexicons of opinion.

3 Towards a specialized vocabulary of opinion related to land-use planning

Our aim is to automatically identify opinion in texts related to a domain application. The project SENTERRITOIRE refers to the territory of the Thau lagoon described in our corpus. Classical methods of opinion mining fail when applied on this corpus as the traditional lexicons are not appropriate to the associated domain (See Section 4.2, Table 1). We also tested a "bag of words" approach by applying classical classification algorithms for supervised learning. Unfortunately, due to the low proportion of data training and the diversity of topics covered, the results were also unsatisfactory (between 50-55% well-classified). We have thus defined the new and generic approach OPILAND which is detailed in the following subsections.

3.1 Overview

At a first step, there is no need for an annotated corpus contrary to conventional methods of opinion mining [18]. However, for assessment purpose, part of the corpus has been annotated by an expert (with positive and negative polarity). To extract opinion from texts, we propose to define a specialized vocabulary of opinions in three steps (See Figure 1): (1) *General vocabulary of pivot opinion (GVPO)*: list of generalized and polarized pivot words (i.e. words extracted from the corpus and existing in at least one of the traditional lexicons of opinion used); (2) *Contextualized vocabulary of opinion (CVO)*: list of words polarized with respect to their context in documents; (3) *Specialized vocabulary of opinion (SVO)*: list of polarized words related to the domain.

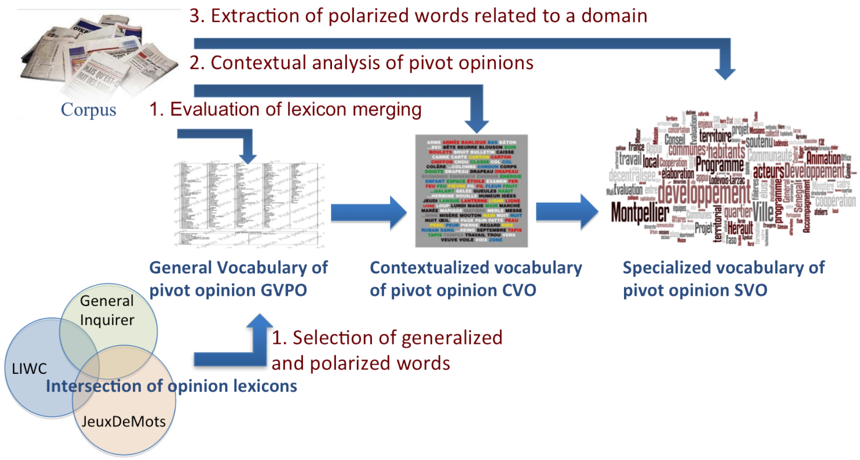

Fig. 1. Opiland Approach

In the following subsections, in order to illustrate all steps involved in the Opiland approach, we will use 100 extracts of documents (40 are negative and 60 are positive) related to land-use planning in the vicinity of Sète (See Figure 2). The extracts 1, 2 and 4 are positive and the extracts 3 and 100 are negative.

The first stage consists in automatically extracting pivot opinions such as: "*magnificent*", "*beautiful*", "*increasing*", "*solution*", "*criticized*", "*fears*", "*leisure*" and "*protect*". To this end, we propose a NLP extraction process built on the basis of opinion lexicons produced by the scientific community. A "Pivot Opinion" PO is a word extracted from the corpus existing in at least one of the three lexicons of opinion used. The second stage involves extracting words which contain pivot opinions in their context (CVO) such as, for example, the words "find" which is located in the same context of the positive pivot word "*solution*" and "*heavily*", located in the same context of the negative pivot opinion "*criticized*". The third stage resides in semi-automatically define a specialized vocabulary of opinion (SVO) related to the domain such as the word "*natural*". Therefore, we propose to adapt a text mining method to highlight a specific vocabulary of opinion related to our domain using the contextualized vocabulary of opinion (CVO).

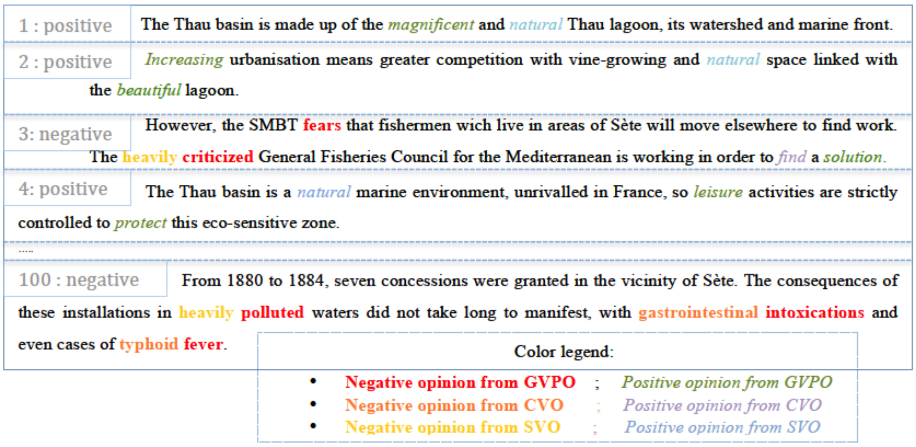

Fig. 2. Document extracts related to the vicinity of XXX

3.2 Construction of the specialized vocabulary of opinion

General vocabulary of pivot opinion

Opinion lexicon are often created manually, semi-automatically, or in a contributive way. In our approach, we use and evaluate three French lexicons:

- Lexicon 1: The General Inquirer lexicon in French [1], is a translated version of the General Inquirer4 which contains syntactic, semantic and pragmatic information about a list of polarized words. For each word, the polarity indicates that a word is positive or negative. This list is available in French after translation, stemming and validation done by two judges [1]. Finally, the lexicon contains 1246 positive words and 1527 negative words.
- Lexicon 2: The LIWC lexicon in French [14] is the translation from the English lexicon Linguistic Inquiry and Word Count5 (LIWC). The words are not in stemming form. In addition, verb conjugations, noun and adjective in sections increase the size of the lexicon (13626 polarized words) . There are also words describing positive and negative emotions.
- Lexicon 3: The lexicon JeuxDeMots [11] is a French lexicon extended to all part-of-speech (noun, verb, adjective, and adverb), and also to a large number of named entities (people, places, brands, events). The lexicon is composed of more than 250000 words obtained with the serious game JeuxDeMots[5]. Moreover an associated system, called LikeIt6, catches polarity information given by users. Currently, 27529 words have been polarized. We consider JeuxDeMots as the other lexicons : we define a value of 1 for positive words and -1 for negative ones.

862 words are ambiguous, i.e. present in one or more selected lexicons in both the positive and negative portion. So we decide to delete them. Our challenge is to find the best combination to obtain a relevant list of linguistic features for opinion called Pivot Opinion *PO*. We firstly propose to merge these different lexicons in a General Vocabulary of Pivot Opinion (GVPO):

[5] http://www2.lirmm.fr/ lafourcade/JDM-LEXICALNET-FR/?C=M;O=D

- $GeneralInquirer \cap LIWC \cap JeuxDeMots = S_1$;
- $GeneralInquirer \cap LIWC = S_2$;
- $GeneralInquirer \cap JeuxDeMots = S_3$;
- $LIWC \cap JeuxDeMots = S_4$;
- $GeneralInquirer = S_5$: words appearing only in the GeneralInquirer lexicon;
- $LIWC = S_6$: words appearing only in the LIWC lexicon;
- $JeuxDeMots = S_7$: words appearing only in the JeuxDeMots lexicon.

Three types of reliability scores are defined: a high score (S_1) to the words contained in the three lexicons, an average score (S_1, S_2, S_3) to the words included in two lexicons, and a low score (S_5, S_6, S_7) to the words contained in only one lexicon[6]. These scores are compared in Section 4.2. From the document extracts presented in Figure 2, the following words are *PO*, i.e., present in at least one of the three lexicons of opinion used: "magnificent, beautiful, leisure, increasing; solution, intoxication, protect, polluted, fever ".

Contextualized vocabulary of pivot opinion
The words contained in the GVPO vocabulary are therefore Pivot Opinions *PO*. Next, other words located in the same context of these *PO* are polarized. In practice, if a word is close to a *PO* (i.e. in a window size, for a given sentence), it is polarized depending on where the *PO* is. The assigned polarity score is defined in formula 1 in which d is the position (in number of words) of the current word W relative to a *PO*.

$$NeighborWordScore(W) = \frac{\sum \frac{WordScore(PO)}{d}}{\sum PO} \quad (1)$$

For each non-pivot word *CW*, all neighboring Pivot Opinions *PO* are identified (i.e. in a window of neighboring words). Next, for each *PO* selected, a *NeighborPS* polarity score is calculated by dividing its polarity score by the distance to *CW*. The *NeighborWordScore* score of *CW* is then the average of its *NeighborPS* score. The CVO vocabulary obtained at this step consists of *PO* from GVPO and other words which are in the same context as *PO*. We use the following concrete example in Figure 3 to illustrate this step. Concerning the word "*heavily* ", if we choose a window of four neighboring words, only the *PO* named "*criticized* " is selected. The *NeighborWordScore* score of "*heavily* " is equal to -1, indicating that this word is negative in this context.

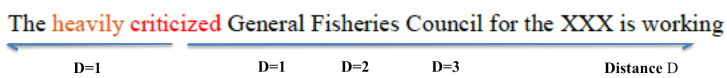

Fig. 3. Syntagm from a document extract related to the vicinity of XXX

[6] These lexicons are available here: http://ekergosien.free.fr/file/SenterritoireProject_S1toS7Lexicons.zip

From the document extracts presented in Figure 2, the following words are some of the words added in the CVO: "natural, find, heavily, gastrointestinal ".

Specialized vocabulary of opinion
Text mining methods allow us to highlight a specific vocabulary of opinion related to our domain. We have implemented a module for the extraction of relevant linguistic features from a corpus based on the contextualized vocabulary of opinion CVO defined in Section 3.2. For each opinion feature O present in the CVO, the number of positive (*nbPos*) and negative (*nbNeg*) documents in which it occurs is counted. A first selection criterion is used to remove features (low presence with respect to *nbTDocs*, i.e, the corpus size) (See Formula 2). We have empirically tested the commonly measures (support, natural logarithm, logarithm to the base 10, tf-idf, etc.) to filter features on the basis of our corpora, and we have chosen to use the measure logarithm to the base 10. Therefore, our results indicate that this measure is a less restrictive than the others when used on small or medium sized corpora.

$$nbPos(O) + nbNeg(O) \leq \log(nbTDocs) \tag{2}$$

From the document extracts presented in Figure 2, the words "*natural* " and "*heavily* " are, for example, retained as candidate for inclusion in the SVO using this selection criterion as shown below:

$$\underbrace{\begin{array}{c} natural \\ ------ \\ nbPos\ = 3 \\ nbNeg\ = 0 \\ log(100) = 2 \end{array}}_{\mathbf{3+0 \geq 2}} \quad \underbrace{\begin{array}{c} heavily \\ ------ \\ nbPos\ = 0 \\ nbNeg\ = 2 \\ log(100) = 2 \end{array}}_{\mathbf{2+0 \geq 2}} \qquad \underbrace{\begin{array}{c} leisure \\ ------ \\ nbPos\ = 1 \\ nbNeg\ = 0 \\ log(100) = 2 \end{array}}_{\mathbf{1+0 \leq 2}}$$

In contrast, the word "*leisure*" is not retained.

We assign to the remaining features O a weighting score WScore (See Formula 3) based on their discriminating factor and the proportion of positive and negative documents in the corpus, *nbTDocNeg* and *nbTDocPos* being the total number of negative and positive documents, respectively. The function *max* with the second parameter *1* is used to avoid that the denominator or the numerator is equal to 0, i.e. the feature O is not present positive or negative documents.

$$WScore(O) = \frac{max(nbPos(O), 1)}{max(nbNeg(O), 1)} \times \frac{nbTDocNeg}{nbTDocPos} \times nbTDocs \tag{3}$$

From the document extracts presented in Figure 2, the WScore for the example "*natural*" would be:

$$WScore(natural) = \frac{max(3, 1)}{max(0, 1)} \times \frac{40}{60} \times 100 = \frac{3}{1} \times \frac{200}{3} = 200$$

In the same way, the WScore for the word "*heavily*" would be 100/3:

$$WScore(heavily) = \frac{max(0,1)}{max(2,1)} \times \frac{200}{3} = \frac{1}{2} \times \frac{200}{3} \approx 33$$

This method is used to process an unbalanced corpus in terms of the number of polarized documents. A feature is deemed to be representative if its occurrence in a document class (positive or negative) is more significant than in the other. Therefore, ambiguous features, which are not representative of a class, are removed. For the remaining features O, we define a representativeness score RS related to the represented class (See Formula 4). The feature is assigned to a positive polarity if its weighting value is greater than, or equal to, T_r, and negative if its weighting value is lower than, or equal to, 1-T_r.

$$\begin{aligned} if\ WScore(O) \geq T_r \times nbTDocs\ Then \\ RS_{pos}(O) = WScore(O) \\ if\ WScore(O) \leq (1 - T_r) \times nbTDocs\ Then \\ RS_{neg}(O) = 1 - WScore(O) \end{aligned} \tag{4}$$

The threshold used in our experiments is $T_r = 65\%$, indicating that all features that do not have at least 65% of the distribution in one of the two classes are removed. Manual validation of these features generates a specialized vocabulary (SVO). For example, the word "*natural*" has been associated with the specialized vocabulary of positive elements related to land-use planning. Indeed, as shown below, the RS score is positive.

$$WScore(natural) \geq 0,65 \times 100\ Then\ RS_{pos}(O)\ =\ 200$$

In contrast, the RS score of the opinion "*heavily*" is assigned as negative:

$$WScore(heavily)\ \leq\ 0,35 \times 100\ Then\ RS_{neg}(O)\ =\ 35$$

3.3 Assigning opinion score to documents

Once the different vocabularies are defined, the next step assigns a polarity score to each document. Two types of preprocessing are previously performed: (1) Statistical preprocessing: Removing not discriminant words based on IDF score (Inverse Documents Frequency), (2) Linguistic preprocessing: A final preprocessing effort consists in weighting pivot words according to their part-of-speech category using Tree-Tagger (i.e. *Cat*).

Based on parameters and preprocessing, an overall polarity score is assigned to each textual object (e.g. word, sentence, and document). Firstly, the Opinions (O) score is calculated according to their presence in the positive lexicon or the negative one (See formula 5, with *polarity (O)* $\in$ {-1, 1}). Thereafter, this polarity is weighted using a reliability score S_i (See Section 3.2).

$$WordScore(O) = S_i \times polarity(O) \tag{5}$$

From the document extract number 2 presented in Figure 2, the WordScore of the opinions "*Increasing*" and "*beautiful*" are both equals to 1 because they are extracted from positive lexicons. Then, the word "*natural*", contained in the CVO has a WordScore equal to 1/3 (See Section 3.2 for the NeighborWordScore description). In this simple example, S_i=1. The score of a sentence S is obtained regarding to the scores assigned to each opinion O (See formula 6). A weight is assigned to the words according to their part-of-speech (i.e. *Cat* parameter). Here, the Cat parameter gives different weights to words according to their grammatical categories (e.g. adjectives, nouns).

$$SentenceScore(S) = \frac{\sum Cat \times WordScore(O)}{\sum Cat \times O} \tag{6}$$

From the document extract number 2 presented in Figure 2, the SentenceScore is about 2,33/3, indicating that it is a positive sentence. In this simple example, Cat=1. Finally, the overall score of the analyzed document is defined as the average of the scores of its associated sentences. In order to affect a score to document by using contextual information (See Section3.2), each word of the context is taken into account (See formula 6).

4 Experiments

4.1 Description of the corpus

We selected newspapers related to land-use planning of a specific lagoon in Thau Agglomration. The corpus consists of 100 documents, called SENT_100. It is divided into two classes of opinion: positive and negative. The opinion is related to the formation of a new metropolitan area gathering several cities. The corpus was validated by geographer experts. It does not contain (i) ambiguous texts presenting different opinions, and (ii) texts without polarity information. The next section presents experimental results with the OPILAND approach on SENT_100.

4.2 Results

Which lexicon? The classification score of polarized document is obtained using the three selected lexicons (i.e. *General Inquirer*, *LIWC*, and *JeuxDeMots*, see Table 1). We note that the *General Inquire* lexicon, more complete than *LIWC*, gives better results (i.e. 57,5%). *JeuxDeMots* lexicon sounds less effective. These results rely to the construction of this resource that does not focus on opinion domain.

Which combination? Furthermore, we merge the three lexicons in order to identify a general vocabulary of pivot opinion. To perform the optimal merging method, we vary all reliability scores S_1, S_2, S_3, S_4, S_5, S_6, and S_7 from 0 to 3[7] (See Section 3.2). The results are presented in Table 2. We observe that

[7] A larger amplitude by weighting the scores from 0 to 10 was not experimentally relevant.

Table 1. Overall score from general lexicons of opinion

Lexicons	Scores of correct classification
General Inquirer	57,5%
LIWC	54,5%
JeuxDeMots	51,5%

the scores are better using *GeneralInquirer* and/or *LIWC* in addition with the intersection of the three lexicons. In the experiments, tests are performed by taking into account Contextual Windows CW (See Table 2). The results show that the identification of polarity are improved when we extend the context window with 4 words preceding and following a target (See Test 7, Table 2).

Table 2. Scores with different combinations of lexicons

Test	CW	S_1	S_2	S_3	S_4	S_5	S_6	S_7	Score
1	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**52,5%**
2	0	1	1	1	1	0	0	0	59,6%
3	0	3	0	2	0	1	0	1	54,5%
4	0	3	0	0	2	0	1	1	54,5%
5	0	3	2	0	0	1	1	0	63,6%
6	0	3	2	0	0	1	0	0	64,6%
7	**4**	**3**	**2**	**0**	**0**	**1**	**0**	**0**	**65,6%**

Which context? The fourth step defines a specialized vocabulary of opinion SVO from the vocabulary CVO and representative features identified by cross-validation. Firstly, every noun, verb, and adjective words are candidates to be features. Then, we apply both filters described in Section 3.2 in order to select relevant features. We consider a feature as representative if its occurrence in a polarized document is at least 65% in favor of one polarity (positive or negative). This threshold enables to select the most discriminating features. By selecting the whole 527 features obtained by cross-validation (see formulas 2 and 3), we obtain a vocabulary of opinion that significantly improves the identification of document polarity with a value at 81,8% (See Table 3, test a).

By weighting the features with their representativeness score, we improve significantly the identification of polarity scores with a value at 91,9% (see Table 3, test b). Note that adding linguistic information (by part-of-speech weighting and by taking into account negation) does not improve classification results.

With these proposed parameters of Opiland approach, classification of land-use planning documents is relevant. Now the question is: are these parameters suitable for other corpora?

Table 3. Overall score with representative features

Test	Representativeness Score	Scores of correct classification
a	1	81,8%
b	RS(O)	91,9%

4.3 Is the Opiland approach generic?

To show the genericity of our approach, we tested the Opiland method on three French corpora associated to the DEFT'07 challenge (1) *CorpusP:* 300 anonymous interventions of politicians, (2) *CorpusV:* 994 reviews of video games, (3) *CorpusM:* 3000 reviews related to movies, books, shows, and comics. Similar to results based on land-use planning corpus, General Inquirer lexicon brings better results than the two others. Moreover, the combination of lexicons based on SVO gives best results (see Table 4). Then, these results underline the genericity of Opiland approaches.

Table 4. Experiments of Opiland on three corpora

	GVPO	CVO	SVO
CorpusP	54,0%	55,0%	69,8%
CorpusV	67,3%	68,4%	73,7%
CorpusM	77,6%	78,8%	82,6%

4.4 Opiland approach vs supervised approaches.

In order to gauge our work, we compared our approach with supervised methods. Table 5 presents results obtained using a bag-of-words approach (without stop words) with Naive Bayes and 10-cross-validation. Note that SVM method provides similar results. The results indicate that supervised methods are inefficient for corpus SENT_100 due to its specificity, the complexity of the used vocabulary, and its small size. Actually, small data sets are still challenging for classical learning approaches. In that way, dedicated approaches such Opiland are more adapted for this kind of corpus.

5 Conclusions and future work

Opiland approach based on a general vocabulary of opinion combines three traditional lexicons of opinion. Our proposal improves identification of document polarity. Indeed Opiland enables the identification of specialized vocabulary, in

Table 5. Naive Bayes classifier results

SENT_100	CorpusP	CorpusV	CorpusM
51,5%	78,6%	90,5%	88,9%

particular for land-use planning opinion. Experiments show our method has a good behavior for processing of small data sets.

In future work, we plan to experiment OPILAND approach on different kinds of documents such as blogs and websites that contain feelings about territorial planning. Indeed the information regarding the opinion is often insufficient. We plan to detail different types of sentiment [3] on multilingual corpora. For instance, the sentiment model of Hourglass [15] is based on four independent dimensions representing the emotional state of the mind (i.e., Sensitivity, Aptitude, Attention, Pleasantness). Each of the four affective dimensions is characterized by six levels which determine the intensity of the expressed/perceived emotion. Note that this model enables the different affective sentiments to co-exist as compound emotions (e.g., love and aggressiveness).

Acknowledgments

The authors thank Pierre Maurel (IRSTEA, UMR TETIS) for his expertise on the corpus. This work was partially funded by the labex NUMEV and the Maison des Sciences de l'Homme de Montpellier (MSH-M).

References

1. Y. Bestgen. Building affective lexicons from specific corpora for automatic sentiment analysis. In *Proceedings of LREC*, pages 496–500, Trento, Italy, 2008.
2. P. Buléon and G. D. Méo. *L'espace social.* Armand Colin, Annales de la recherche urbaine, 2005.
3. E. Cambria, C. Havasi, and A. Hussain. Senticnet 2: A semantic and affective resource for opinion mining and sentiment analysis. In *FLAIRS Conference.* AAAI Press, 2012.
4. B. Debarbieux and M. Vanier. *Ces territorialités qui se dessinent.* Editions de l'Aube, 267 pages, Datar, 2002.
5. B. Duthil, F. Trousset, M. Roche, G. Dray, M. Plantié, J. Montmain, and P. Poncelet. Towards an automatic characterization of criteria. In *International Conference on Database and Expert Systems Applications (DEXA'11)*, volume 1, pages 457–465, Toulouse, France, 2011. Springer-Verlag, LNCS.
6. A. Esuli and F. Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *5th Conference on Language Resources and Evaluation*, pages 417–422, 2006.
7. W. Fan, S. Sun, and G. Song. Sentiment classification for chinese netnews comments based on multiple classifiers integration. In *Proc. of the Int. Joint Conf. on Comp. Sciences and Optimization*, pages 829–834, 2011.

8. A. Joshi, P. Balamurali, P. Bhattacharyya, and R. Mohanty. C-feel-it: a sentiment analyzer for microblogs. In *Proc. of HLT*, pages 127–132, 2011.
9. A. Kennedy and D. Inkpen. Sentiment classification of movie reviews using contextual valence shifters. In *Computational Intelligence*, volume 22(2), pages 110–125, 2006.
10. B. Klebanov, E. Beigman, and D. Diermeier. Vocabulary choice as an indicator of perspective. In *Proceedings of the ACL 2010 Conference Short Papers*, ACLShort '10, pages 253–257, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.
11. M. Lafourcade. Making people play for lexical acquisition. In *Proc. 7th Symposium on Natural Language Processing (SNLP 2007)*, pages 13–15, 2007.
12. B. Liu. Sentiment analysis and opinion mining. page 167. Morgan and Claypool Publishers, 2012.
13. A. Pak and P. Paroubek. Microblogging for micro sentiment analysis and opinion mining. In *TAL*, volume 51(3), pages 75–100, 2010.
14. A. Piolat, R. Booth, C. Chung, M. Davids, and J. Pennebaker. La version française du dictionnaire pour le liwc: modalités de construction et exemples d'utilisation. In *Psychologie Française*, volume 56(3), pages 145–159, 2011.
15. R. Plutchik. The nature of emotions. *American Scientist*, 89(4):344–350, 2001.
16. D. R. Rice and C. Zorn. Corpus-based dictionaries for sentiment analysis of specialized vocabularies. In *Proceedings of NDATAD 2013: New Directions in Analyzing Text as Data Workshop 2013*, London, England, 2013.
17. M. P. M. Roche, G. Dray, and P. Poncelet. Is a voting approach accurate for opinion mining? In *Proc. of DatawWrehousing and Knwoledge discovery (DaWaK'08)*, pages 413–422, 2008.
18. J. Torres-Moreno, M. El-Beze, F. Bechet, and N. Camelin. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proc. of ACL*, pages 417–424, 2009.
19. P. Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proc. of ACL*, pages 417–424, 2002.
20. M. Vanier2. Territoires, territorialité, territorialisation - controverses et perspectives. In *PUR*, pages 417–424, 2002.
21. J. Wiebe and E. Riloff. Finding mutual benefit between subjectivity analysis and information extraction. In *IEEE Transactions on Affective Computing*, volume 2 (4), pages 175–191, 2011.