Extracting Knowledge from Collaboratively Annotated Ad Video Content

Manolis Maragoudakis, Katia Kermanidis, Spyros Vosinakis

Extracting Knowledge from Collaboratively Annotated Ad Video Content

Manolis Maragoudakis[1], Katia Lida Kermanidis[2], Spyros Vosinakis[3]

[1]Department of Information and Communication Systems Engineering, University of the Aegean, 83200 Karlovasi, Samos, Greece
[2]Department of Informatics, Ionian University, 49100 Corfu, Greece
[3]Department of Product and Systems Design Engineering, University of the Aegean, 84100 Ermoupoli, Syros, Greece
kerman@ionio.gr, mmarag@aegean.gr, spyrosv@aegean.gr

Abstract. Creative advertising support tools have relied so far on static knowledge represented in creativity templates and decision making systems that indirectly impose restrictions on the brainstorming process. PromONTotion is a system under development that aims at creating a support tool for advertisers for the creative process of designing a novel ad campaign that is based on user-driven, generic, automatically mined and thus dynamic semantic knowledge. Semantic terms and concepts are collaboratively provided via crowdsourcing. The present work describes data mining techniques that are applied to the collected annotations and some interesting initial results regarding ad content, ad genre, ad style and ad impact/popularity information.

1 Introduction

Creative advertising is the process of capturing a novel idea for an ad campaign and designing its implementation. As it constitutes one of the highest-budget enterprises today, several studies have been published regarding the impact of advertising (Amos et al., 2008; Aitken et al., 2008), as well as creativity in advertising (Hill and Johnson, 2004).

Several tools have been proposed in the literature for supporting the creative brainstorming process of ad design. A number of creativity support tools have been proposed, to help ad designers come up with novel ideas for setting up a new campaign. They usually focus on using creativity templates (Goldenberg et al., 1999), decision making systems, like ADCAD (Burke et al., 1990), linguistic wording schemata (Blasko and Mokwa, 1986), databases of predefined concepts and their associations, like IdeaFisher (Chen, 1999). Idea Expander (Wang et al., 2010) is a tool for supporting group brainstorming by showing pictures to a conversing group, the selection of which is triggered by the conversation context. Opas (2008) presents a detailed overview of several advertising support tools.

Based on the belief that total freedom is not the most efficient way for enhancing the creative process, but constraining it with the use of a limited number of idea-

adfa, p. 1, 2011.

forming patterns is, these tools force upon the advertiser a certain restricted way of thinking. Concepts, associations are static and predefined, and the user is guided through the problem entities to make connections between them. Required facts and rules are expert-dependent and usually quite hard to craft. Dictionaries and databases are static and non-expandable. According to Opas (2008), static, passive, expert-dependent knowledge models can hurt creativity.

PromONTotion is a creative advertising support tool that relies on generic, automatically acquired knowledge. The core of the tool is comprised of a semantic thesaurus, i.e. an ontology of terms and concepts that are related to the content, style, genre, artistic features, consumer impact and popularity of television ad videos. These terms and concepts are provided via crowdsourcing by players of a multi-player quiz/combat/action style computer game, House of Ads, developed especially for this purpose and described in detail in Kermanidis et al. (2013).

The present work describes in detail the mining experiments performed on an initial set of annotations. Mining techniques are applied to the provided data for the extraction of correlation information, of interdependencies between various features of the ad, of knowledge related to what increases the ad impact and popularity.

PromONTotion relies on very limited predefined knowledge, i.e. the ontology backbone, its hierarchical structure. The ad knowledge provided by the tool is data-driven and automatically derived making PromONTotion generic, dynamic, scalable, expandable, robust and therefore minimally restricting in the creative process and imposing minimal limitations to ideation or brainstorming.

This article is structured as follows: Section 2 presents a sketch of the ontological backbone used to identify key elements of the advertisement domain, while Section 3 describes the data mining approaches used upon collected data, in order to identify meaningful patterns and associations between objective ad parameters and subjective user feedback. Finally, the paper concludes in section 4.

2 Modeling The TV Ad Domain

The domain model in this work is represented as a hierarchical ontology of TV ad video concepts and terms. The ontological backbone structure is the only piece of knowledge that is manually crafted, by advertising domain experts. It contains terms and concepts that are related to the video content, the ad style and genre, its production values, its artistic features, as well as its impact and its impression print on consumers. The structure is scalable. Some representative product and services types are included currently, but the list is easily expandable. In detail, the ontology structure and categories are shown in Figure 1.

1. Artistic features
   (a) Sound
      (i) Music/Song recognisability
      (ii) Song Music type
   (b) Filming
      (i) Photography
      (ii) Style
2. Location
   (a) Indoors

- (b) Outdoors
3. Ad impact
   - (a) Convincing power
   - (b) Opinion
   - (c) Improvement suggestions
4. Production
   - (a) Producer
   - (b) Director
   - (c) Production quality
5. Participating elements
   - (a) Main character
     - (i) Recognisability
     - (ii) Type
       - (1) Human
         - (a) Gender
         - (b) Age
         - (c) Occupation
       - (2) Animal
       - (3) Inanimate
   - (b) Key participants
6. Message communication
   - (a) Ad structure
   - (b) Linguistic schemata
   - (c) Indirect Critique on competition
   - (d) Humoristic elements
   - (e) Tag lines
   - (f) Brand name
7. Product type
   - (a) Product
     - (i) Food
     - (ii) Beverage
     - (iii) Electric device
       - (1) Device type
       - (2) Energy class
     - (iv) Electronic device
     - (v) Store
     - (vi) Vehicle
       - (1) Type
       - (2) Value
     - (vii) Household
   - (b) Service
     - (i) Telecommunications
     - (ii) TV
     - (iii) Banking
     - (iv) Insurance
     - (v) Healthcare
8. Target group
9. Product origin

Fig. 1. The Ontology backbone.

Following the trend of serious games paved by von Ahn (2006) and game design for ontology populating (Siorpaes and Hepp, 2008), the *House of Ads* multiplayer browser-based action game (Kermanidis et al., 2013) was designed and developed for collaboratively providing ad content annotations, i.e. for ontology populating.

3 Mining Through Advertisement Terms

An initial set of annotations of 74 videos were used for performing preliminary mining experiments. The data were provided by 21 players. For the preliminary experiments 20 different ad videos were selected and every player was assigned the annotation of 3 to 5 videos.

Exploratory as well as predictive data mining is applied to the collected data, for detecting patterns and interdependencies among ad features, as well as their association with the ad's impact on consumers/players and its popularity. All categories proposed by the domain experts are considered features (41 in number) that define the ad domain space, and every ad video constitutes a data instance. It needs to be noted that 13 out of the 41 features present a missing value percentage of more than 80%, due to the hierarchical structure of the ontology. The input of lower-level annotations depends on the corresponding value of the parent node in the hierarchy. If the specific parent value is not provided, values of the sub-nodes are missing.

3.1 Mining Associations

The Apriori algorithm (Liu et al., 1998) on the Weka[1] machine learning workbench was used for extracting association rules from the data. Given the small number of distinct videos involved in the preliminary experiments, rules concerning only the objective (content/style/genre) data regarding the ad, e.g.

```
Service=telecom=>GenderOfMainCharacter=male(Confidence=1)
TargetGroup=women=>Location=inside(Confidence=1)
```

have not been given attention. Although they are interesting. the small number of videos does not allow for their generalization. Rules related to the subjective data, however, reveal interesting findings. For example, they show an absolute correlation between the convincing ability of the ad and how the player liked the ad:

```
Opinion=LikeNo => Convincing=No (Confidence=1)
Opinion=LikeALittle => Convincing=So-So (Confidence=1)
Opinion=LikeALot => Convincing=ALot (Confidence=1)
```

A strong correlation between these two features was expected, but the absolute correlation constitutes an interesting finding. These rules emerged after removing attributes (e.g. ProductType) that led to trivial, uninteresting rules, like

```
Service=telecom => ProductType=service
```

3.2 Classification

Predictive experiments were run on the Weka platform for identifying the correlation of the various ad features with the ad's impact on the players. To this end the *Opinion* feature denoting how the player liked the ad (with values *No, ALittle, Considerably, ALot*), the *Convincing* feature, indicating how persuaded the player felt by the ad (with values *No, SoAndSo, ALot*), and the *Changes* feature, that lists the ad features that the player does not like and would change if he could (with values *Everything, Nothing, TheCharacters, TheLocation, TheContent, TheMusic, TheGenre, Other*), were used as class labels. The main goal of this task is to focus on exploratory data analysis and reveal interesting interdependencies. As the following experiments show, the data is too few, the percentage of missing values too high and the feature space too large to expect high accuracy predictive performance at this phase.

[1] http://www.cs.waikato.ac.nz/ml/weka/

Due to their comprehensibility and their ability to cope with missing values using distribution-based imputation (Saar-Tsechansky and Provost, 2007), decision trees are the first algorithm employed for inducing a classification model. Support Vector Machines (Cortes and Vapnik, 1995) were also experimented with, as they have been proven to perform well on problems with few learning items in large feature spaces. The Sequential Minimal Optimization algorithm (Platt, 1998) is used for training the SVM classifier and a third degree polynomial kernel function is chosen. Tables 1, 2 and 3show the results for the three class attributes and the two classification algorithms. All experiments were run using 10-fold cross validation.

For the Changes attribute, the results for the remaining values are zero, or close to zero, due to their sparse occurrence in the data. In the last line of Table 3, the average precision and recall value over all class values is shown instead.

***Opinion* Attribute**	**C4.5**		**SVMs**	
	Precision	Recall	Precision	Recall
No	0.389	0.467	0.462	0.4
ALittle	0.524	0.55	0.464	0.65
Considerably	0.35	0.412	0.429	0.353
ALot	0.8	0.545	0.632	0.545

Table 1. Results for the *Opinion* class.

***Convincing* Attribute**	**C4.5**		**SVMs**	
	Precision	Recall	Precision	Recall
No	0.579	0.524	0.632	0.571
SoAndSo	0.718	0.7	0.727	0.8
ALot	0.5	0.615	0.636	0.538

Table 2. Results for the *Convincing* class.

***Changes* Attribute**	**C4.5**		**SVMs**	
	Precision	Recall	Precision	Recall
Everything	0.357	0.455	0.3	0.273
Nothing	0.676	0.758	0.5	0.848
Content	0.364	0.364	0	0
Average (all classes)	0.449	0.5	0.268	0.419

Table 3. Results for the *Changes* class.

Keeping in mind that the available learning items in these preliminary experiments are very few, the number of features is significant, the percentage of missing values is very high and the classification task is not binary, these results at this stage are quite promising for the experiments to be conducted when the annotation phase has progressed and learning items have accumulated.

The ‘simple’ SVMs classifier used is in most cases outperformed by C4.5. This is largely attributed to the missing values problem. It has been reported in the literature

(Pelckmans et al., 2005) that SVMs require additive models (e.g. component wise SVMs and LS-SVMs) to cope adequately with missing values.

The attributes that are significant for predicting the convincing power of the ad, included in the decision tree model, include (except the opinion of the player) the fame of the main character (famous characters seem to be more convincing) and the genre of the ad (cartoon ads seem to imprint their message more efficiently on consumers compared to realistic genre ads). Regarding the prediction of the player's opinion of the ad, except for the convincing attribute, the location of the ad story (urban environment, home, country), as well as the artistic features (e.g. music recognition) play an important role. Finally, regarding the prediction of changes the players propose, the opinion is understandably the most important feature, with a negative correlation. The use of taglines and famous logos is also strongly correlated to the prediction of the content value of changes, possibly indicating that they are a communication tool consumers get easily tired of.

3.3 Bayes Simulation

For the purpose of our study, Bayesian Networks (BN) are used for modelling of a marketing process. More specifically, subjective advertisement attributes, such as User Opinion, Convincing Level and Changes Required are modelled in parallel with objective attributes, as thoroughly described in Section 2. Unlike classification analysis, BN modelling can support two semantic operations, namely diagnostic inference (i.e. from effects to causes) and causal inference (i.e. from causes to effects) (Friedman and Goldszmidt, 1996). As regards the former operation, we are mostly interested in providing the most probable value for the subjective attributes, given the values of remaining objective ones and benchmark its performance against the well-known classification algorithms. Concerning the latter operation, the aim is to model the dependency between input and output parameters, and based on the extracted modelling outcome, calculate the objective input parameters that guarantee an optimal subjective output (i.e. with a maximal probability). The aforementioned task can be accomplished by entering the intended output as evidence and then calculate the most probable instantiation of the remaining input parameters that caused the intended output.

3.3.1. Diagnostic Inference

In this setting, a BN was built from the initial dataset, using the K2 learning approach (Cooper and Herskovits, 1992). Upon, learning, the BN was used in order to generate data instances. The generation process is as follows: a node without parents is selected as a starting point. According to its conditional probability table (CPT), a value is drawn using a roulette wheel strategy. This value is passed to the child node and again, using the CPT from the BN, a value for that child node is also selected. This process is progressing to all nodes in the BN until a sample containing values for all attributes is generated. The new, augmented dataset was evaluated using a variation of Naïve Bayes classifier which incorporates elements from Bayesian Networks, namely a variation of Tree-Augmented Naïve Bayes (TAN) algorithm (Friedman et al., 1997). In our approach, the class node is kept separately from the other, attribute nodes. A BN is learned and then the class node is placed as the parent of all other nodes in order to favor classification. For comparative analysis, tables 4, 5 and 6 show classification results of TAN against C4.5 and SVMs (parameter setup identical to the previous experiments) with the simulated dataset.

***Opinion* Attribute**	**C4.5**		**SVMs**		**TAN**	
	Precision	Recall	Precision	Recall	Precision	Recall
No	0.553	0.638	0.556	0.542	0.545	0.602
ALittle	0.452	0.207	0.328	0.321	0.423	0.342
Considerably	0.386	0.686	0.354	0.329	0.401	0.312
ALot	0.808	0.619	0.569	0.623	0.762	0.687

Table 4. Classification results with the simulated dataset for the *Opinion* class.

***Convincing* Attribute**	**C4.5**		**SVMs**		**TAN**	
	Precision	Recall	Precision	Recall	Precision	Recall
No	0.691	0.732	0.649	0.626	0.661	0.673
SoAndSo	0.797	0.723	0.689	0.711	0.691	0.715
ALot	0.665	0.784	0.6	0.577	0.701	0.873

Table 5. Classification results with the simulated dataset for the *Convincing* class.

***Changes* Attribute**	**C4.5**		**SVMs**		**TAN**	
	Precision	Recall	Precision	Recall	Precision	Recall
Everything	0	0	0.153	0.182	0.134	0.152
Nothing	0.436	1	0.447	0.522	0.356	0.466
Content	0	0	0.172	0.181	0.231	0.199
Average (all classes)	0.19	0.436	0.259	0.292	0.232	0.243

Table 6. Classification results with the simulated dataset for the *Changes* class.

The TAN classifier is considered superior than Naïve Bayes (Friedman et al., 1997), mainly because of the fact that input attributes are not considered as independent among each other, given the class. On the contrary, the input attributes in a TAN representation form a BN that could explain possible inter-relations found in data.

From the above results, it is clear that *Convincing*, when used as the class, could provide the most satisfactory outcomes, compared with the other two attributes, *Opinion* and *Changes*. The TAN methodology performs slightly better than SVM and is very close to C4.5, although in some cases TAN seem to outperform both learners.

3.3.2. Causal Inference

The aim of this experiment is to model the dependency between input and output parameters, and based on the extracted modeling outcome, to calculate the input parameters that guarantee an optimal output with a maximal probability. The structure of the BN allows for such an operation by entering target values on selected nodes and then estimating the most probable instantiation of the remaining input parameters that provides the entered output.

The exact scenario is as follows: a domain expert is interested in observing the effect that *Opinion* and *Convincing* have to the other variables, given that *Changes* is known. The first scenario (optimistic) assumes that *Opinion*=***Alot*** and *Convincing*=***Alot*** given Changes=**nothing**, while the second scenario (pessimistic) considers the hypothesis that *Opinion*=***No*** and *Convincing*=***No*** given Changes=**everything.**

By measuring the progression of the values for each input variable given the scenario assumptions, one can draw significant conclusions on the effect of each parameter. For reasons of readability, only the variables that belong to the Markov Blanket of the nodes *Opinion* and *Convincing* are shown in Tables 7 and 8. The Markov Blanket of a node contains the set of parent, children and children's parent of that node. It is proven that any node that is not within the Markov Blanket, is not influencing or not being influenced by that node. Table 7 tabulates the most probable attribute instantiations for the input attributes, given evidence described in scenario 1, while Table 8 shows the same information given the evidence representing the pessimistic scenario.

Taking a closer look at the first scenario, one could observe that when a user has a positive opinion and states that an advertisement is convincing, the most probable product being advertised is a drink. Moreover, the main character is human, middle aged and not a famous actor or celebrity. Since the most probable product is a drink, the advertisement does not promote something novel and moreover, the location is indoor and therefore the photography is neither picturesque nor landscape. Finally, the advertisement is an isolated episode and there are no famous utterances used.

Attribute	Value
Product	Drink
Genre	Realistic
MainCharacter	Human
Fame	Everyday character
Age	MiddleAged
Occupation	Employee
Humor	noHumor
SongRecognizable	Yes
Novelty	Known product/service
Location	Indoor
Photography	Other
Structure	Isolated episode
FamousLines	No famous lines

Table 7. The most probable attribute instantiations for the optimistic scenario. (i.e. *Opinion*=***Alot*** and *Convincing*=***Alot*** given Changes=**nothing)**

On the contrary, using the pessimistic scenario, in which a user holds a negative opinion and is not persuaded by the advertisement, one could observe that the product is not listed (note: the listed products were: electronic services, clothes, drink, electronic device, household, vehicle and store). An interesting remark is that despite what most would assume, the presence of famous persons, young in age and the inclusion of a humoristic scenario do not guarantee user satisfaction in terms of convincement and opinion. On the other hand, these instantiations seem to cause a negative impact. Finally, it is noteworthy that in a pessimistic scenario, the location is a workplace.

Attribute	Value
Product	Other
Genre	Other
MainCharacter	Human
Fame	Famous person
Age	Young
Occupation	Other
Humor	Humoristic scenario
SongRecognizable	No
Novelty	New product/service
Location	Workplace
Photography	Notvalid
Structure	Isolated episode
FamousLines	No famous lines

Table 8. The most probable attribute instantiations for the pessimistic scenario. (i.e. *Opinion*=***No*** and *Convincing*=***No*** given Changes=**everything)**

4 Conclusion

In this paper we described the results of some preliminary mining experiments that were performed on an initial set of ad content annotations. More specifically, Association Discovery, Classification and Bayesian network analysis were incorporated in order to examine potentially meaningful patterns. The major obstacle that the limited amount of training data was confronted by training a Bayesian network and then applying a simulation step that generated a plethora. Experiments with the initial and the augmented dataset using various classifiers showed that classification accuracy may be improved using Bayesian modelling. Moreover, causal inference was also applied using the BN structure, simulating two different scenario settings, which revealed some useful insight on input attributes and their impact on subjective attributes such as User opinion and Convince level.

More interesting and confident findings will be revealed after processing an annotation set of significant size, when the annotation process of PromONTotion has progressed. The use of mining algorithms that are better suited to cope with data consisting of a significant percentage of missing values is also planned for future research.

Acknowledgements

This Project is funded by the National Strategic Reference Framework (NSRF) 2007-2013: ARCHIMEDES III – Enhancement of research groups in the Technological Education Institutes.

References

R. Aitken, B. Gray and R. Lawson. 2008. Advertising Effectiveness from a Consumer Perspective. International Journal of Advertising, 27 (2): 279–297.

C. Amos, G. Holmes and D. Strutton. 2008. Exploring the Relationship between Celebrity Endorser Effects and Advertising Effectiveness. International Journal of Advertising, 27 (2): 209–234.

V. Blasko, M. Mokwa. 1986. Creativity in Advertising: A Janusian Perspective. Journal of Advertising, 15(4): 43–50.

R. Burke, A. Rangaswamy, J.Wind and J. Eliashberg. 1990. A Knowledge-based System for Advertising Design. Marketing Science, 9(3): 212–229.

Z. Chen. 1999. Computational Intelligence for Decision Support. CRC Press, Florida.

G.F. Cooper, and E. Herskovits, A Bayesian Method for the Induction of Probabilistic Networks from Data, Machine Learning, Vol. 9, 1992.

C. Cortes, V. Vapnik. 1995. Support-vector Networks. Machine Learning, 20(3): 273-297.

N. Friedman, and M. Goldszmidt, “Building Classifiers and Bayesian Networks”, Proceedings of the National Conference on Artificial Intelligence, AAAI Press, Menlo Park, California, 1996.

N. Friedman, D. Geiger and M. Goldszmidt, “Bayesian Network Classifiers”, Machine Learning, 29, 131–163, 1997.

J. Goldenberg, D. Mazursky and S. Solomon. 1999. The Fundamental Templates of Quality Ads. Marketing Science, 18(3): 333–351.

R. Hill, L. Johnson. 2004. Understanding Creative Service: A Qualitative Study of the Advertising Problem De-lineation, Communication and Response (APDCR) Process. International Journal of Advertising, 23(3): 285–308.

K. Kermanidis, M. Maragoudakis, S. Vosinakis and N. Exadaktylos. 2013. Designing a Support Tool for Creative Advertising by Mining Collaboratively Tagged Ad Video Content: The Architecture of PromONTotion. Proceedings of the International Conference on Artificial Intelligence Applications and Innovations (AIAI). Paphos, Cyprus.

M. Lee, H. Shen, J. Z. Huang and J. S. Marron. 2010. Biclustering Via Sparse Singular Value Decomposition. Biometrics, 66: 1087-1095.

B. Liu, W. Hsu, Y. Ma. 1998. Integrating Classification and Association Rule Mining. Fourth International Conference on Knowledge Discovery and Data Mining, pp. 80-86.

K. MacCrimmon, C. Wagner. 1994. Stimulating Ideas Through Creativity Software. Management Science, 40(11): 1514–1532.

T. Opas. 2008. An Investigation into the Development of a Creativity Support Tool for Advertising. PhD Thesis. Auckland University of Technology.

K. Pelckmans, J. de Brabanter, J. A. K. Suykens and B. de Moor. 2005. Handling Missing Values In Support Vector Machine Classification. Neural Networks, 18(5-6): 684-692. Elsevier.

John Platt. 1998. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. Advances In Kernel Methods-Support Vector Learning.

M. Saar-Tsechansky, F. Provost. 2007. Handling Missing Values When Applying Classification Models. Journal of Machine Learning Research, 8:1625-1657.

K. Siorpaes, M. Hepp. 2008. Games With A Purpose For The Semantic Web. IEEE Intelligent Systems: 1541-1672.

V. Vapnik. 1995. The Nature of Statistical Learning Theory. Springer-Verlag. ISBN 0-387-98780-0.
L. von Ahn. 2006. Games with a Purpose. IEEE Computer, 39(6): 92-94.
H. C. Wang, D. Cosley and S. R. Fussell. 2010. Idea Expander: Supporting Group Brainstorming with Conversa-tionally Triggered Visual Thinking Stimuli. Proceedings of the ACM Conference on Computer Supported Co-operative Work (CSCW). Georgia, USA.