CAUSAL ASSOCIATION MINING FROM GERIATRIC LITERATURE

Anand Krishnan[1] Jon Sligh[2] Eric Tinsley[2] Natalie Crohn[2] Jean Bandos[3] Heather Bush[2] Jason DePasquale[2] Mathew Palakal[1]
[1]Indiana University School of Informatics and Computing Indianapolis, IN 46202
[2]Allegiant Consulting Indianapolis, IN 46290
[3]Marian University Indianapolis, IN 46222

Abstract

Literature pertaining to geriatric care contains rich information regarding the best practices related to geriatric health care issues. The publication domain of geriatric care is small as compared to other health related areas, however, there are over a million articles pertaining to different cases and case interventions capturing best practice outcomes. The knowledge extracted from these articles could be harvested and translated from research to practice in a quicker and more efficient manner. Geriatric literature contains multiple domains that contain information such as interventions, information on care for elderly, case studies and real life scenarios. These articles contain a variety of causal relationships such as the relationship between interventions and disorders. The goal of this study is to identify these causal relations from published abstracts. Natural language processing and statistical methods were adopted to identify and extract these causal relations with a precision of 79.54% and recall of 81%.

Keywords— Geriatric, Causal associations, Semantic tagging, CRF.

I. INTRODUCTION

Geriatric literature is comprised of documents that contain information about Geriatric Syndromes [1]. These syndromes are groups of specific signals and symptoms that occur more often in the elderly and can impact patient morbidity and mortality. These are becoming increasingly important for nurses and care providers to consider as the patient population ages. Numerous publications are available regarding the "best practices" for geriatric care to address Geriatric Syndromes and other geriatric related issues. Though the number of publications specific to geriatric care is small, there are several published peer-reviewed articles that contain different interventions, use-case scenarios, and problems that the elderly face. One way of collecting data is to capture the abstracts that provide a synopsis of what the article contains and apply text mining techniques like Pattern Recognition, Classification, Neural Networks, Support Vector Machines, and Cluster Analysis to extract relevant information from them [2,3,4,5,6].

The experts in the geriatrics (domain experts) chose 42 of the most common Geriatric Syndromes for this work. Table 1 shows the list of all Care Categories identified for this study. This stand-alone system uses a new technique which integrates Syntactic tagging, Semantic tagging and Conditional Random Fields for extraction of causal relations from 2280 Pubmed [7] abstracts.

TABLE 1: CARE CATEGORIES

Fall Risk	Health History	Medical Alerts
Cognition	Well-Being	Care Provision
Medication Management	Safety and Assistive Devices	Caregiver Support
Safety	Supportive Services	Sensory
Providers	Elder Abuse	Medical Issues
Anxiety	Pain Management	Chronic Disease
Sleep	Legal	Spiritual
Financial	Legal Older Adults	Mobility
Information Preference	Activities of Daily Living	Wellness Prevention
Instrumental Activities Of Daily Living	Alternative Living Options	Depressive Symptoms
Nutrition	Substance Abuse	Emotional
Environmental	Insurance Issues	Intellectual
Social	Preferences	Social Interaction
Health Status	Stress Management	Incontinence

II. RELATED WORKS

Causal relations can be expressed in different ways and differ from one domain to the other. The relations can be expressed between two sentences, two phrases, between subject and object noun phrases, in intra-structure of noun phrases and even between paragraphs that describe events.

Natural Language Processing

Natural Language Processing (NLP) is an area of research that explores how natural language text can be understood and manipulated by computers to do useful things [8]. For natural language, syntax provides rules or standardized features to put together words to form components of sentence.

Extracting Causal Associations

Sentences like "*Inflation affects the buying power of the dollar.*", "*Cigarette smoking causes cancer.*", "*Happiness increases with sharing.*" clearly demonstrates a relation between one event or entity (*Inflation, Cigarette, Happiness*) to another (*buying power, cancer, sharing*) with temporal relations like "*affect*", "*causes*" and "*increases*". Extraction of such causal relations from any literature can be very tricky if

978-1-4799-7502-0/14 $31.00 © 2014 IEEE
DOI 10.1109/BIBE.2014.44

we understand the complex nature of natural language. Early research in causal association extraction analysis started with a manually curated causal pattern set to find causal relationships from literature. The causal patterns Khoo et al. [13] investigated an effective cause-effect information extraction system from newspaper using simple computational method. They demonstrated an automatic method for identifying and extracting cause-effect information in text from the Wall Street Journal using linguistic clues and pattern-matching. Marcu et al. [14] hypothesized that lexical item pairs can help in finding discourse relations that hold between the text spans in which the lexical items occur. They used sentence pairs connected with the phrases "*because*" and "*thus*" to distinguish the causal relation from other relations. This method used only nouns, verbs and cue phrases in each sentence/clause. Non-causal lexical pairs were also collected from the sentence pairs to compose the Naive Bayes classifier. Causality extraction problem can be addressed by building a dictionary of causal words extracted from literature.

Causal relation extraction can also be done in a semi-automatic form. The method presented by [15] shows a semi-automatic method of discovering generally applicable lexico-syntactic patterns that refer to the causal relation. They discuss several ways in which a causal relation can be expressed but focus on a single form, *<NounPhrase1 verb NounPhrase2>*. Lexico-syntactic patterns are discovered from a semantic relation for a list of noun-phrases extracted from Wordnet 1.7 [16] and patterns are extracted that links the two selected noun phrases by searching a collection of texts. This gave a list of verb/verbal expressions that refer to causation. Once the list is formed, the noun phrases in the relationship of the form *<NounPhrase1 verb NounPhrase2>* can express explicit or implicit states. Only certain types of such states were considered for the study. These relationships were analyzed and ranked.

Syntactic Tags - Parts-Of-Speech Tagging (POS)

Syntactic features describe the relations between tokens by indicating the functional role of the token. Parts-Of-Speech tagging is a tool used to identify a contextually proper morpho-syntactic description for each ambiguous word in a text [9]. Natural language has several different parts of speech that include nouns, pronouns, verbs, adjectives, adverbs, prepositions, conjunctions and interjections. When a sentence is passed through a tagging process, the natural language text is assigned its parts of speech. There are several POS tagging tools such as Brill Tagger [10] and Stanford POS tagger [11], but Medpost [12] POS tagger has an accuracy of 97% which is one of the most popular tagging tools.

Semantic Tagging

Semantic tagging is a method of assigning tags, symbols or markers to text strings which can help in identifying their meaning so that the string and its meaning can be made discoverable and readable not only by humans but also by computers. It involves annotating a corpus with instructions that specifies various features and qualities of meaning in the corpus [17]. POS taggers are simplest examples of a semantic tagging system. The study in [18] performs "Sense Tagging" which is a process of assigning a particular sense from some vocabulary to the content work in a text. It discusses the approaches that are applied for Word Sense Disambiguation (WSD). The method uses Dictionary look-up module, Parts-of-speech tagging using Brill Tagger and simulated annealing algorithm to optimize the dictionary definition overlap for the remaining sentence. At the end of the algorithm, a single sense is assigned to each token which is the tag associated with that token.

Conditional Random Field

CRFs are undirected graphical models that model the conditional distribution p(x|y) rather than joint probability distribution p(y,x) and trained to maximize the conditional probability of outputs given the inputs[19]. CRFs avoid the label bias problem, which are a weakness shown by Maximum Entropy Markov Models (MEMMs) and other conditional Markov models based on directed graphical models. CRF surpasses the performance of both MEMMs and HMMs on a number of real-world tasks.

A probability distribution of *p(x,y)*, over a set of random variables V=x U y, can be represented by a product of distributions that represent a smaller set of the full variable set [20].

$$(x, y) = \frac{1}{Z}\prod_{a \in F} \Phi_a(x_a, y_a) \qquad (2.1)$$

$$(F = a \subseteq \mathrm{V}) \qquad (2.2)$$

Where, *a* is a subset of V

$$\mathrm{x} = \langle \mathrm{x1}, \mathrm{x2}, \ldots, \mathrm{xn} \rangle \qquad (2.3)$$

is the set of input variables for instance a sequence of tokens and

$$\mathrm{y} = \langle \mathrm{y1}, \mathrm{y2}, .., \mathrm{yn} \rangle \qquad (2.4)$$

$$Z = \sum_{x,y} \prod_{a \in F} \Phi_a(x_a, y_a) \qquad (2.5)$$

is a set of output variables which for our case are the corresponding cause, effect or out tags for the tokens in a sentence. And *Z* defined in Eq. (2.1) is a constant that normalize Eq. (2.5) distribution to one.

Where

$$\Phi_a(x_a, y_a) = exp\{\textstyle\sum_{x,y} \lambda_{ak}\, \mathrm{f}_{ak}(x_a, y_a)\} \qquad (2.6)$$

The weights will be learned in a training procedure to positively reinforce the feature functions that are correlated with the output labels or assign negative values to feature functions that are not correlated with the output labels and zero values to uninformative feature functions. CRF can be used for Signal detection [21] (identifying a data modules) in textual material and sequence labeling of text where there is lack of annotation and multi-domain vocabulary which makes traditional mining techniques ineffective. MALLET [22] provides tools for sequence tagging. It makes use of algorithms

like Hidden-Markov Models [23], Maximum Entropy Markov Model and Conditional Random Fields. To train the CRF model, data is manually annotated to form a training set. A validation set is used to verify the performance of the trained model. [24] trains the CRF model with features as a set of cause and effect tags to identify keywords in sentences. It was able to produce a precision of 84.6% and recall of 87%.

III. Design and Implementation

All NLP systems work on a systematic approach. Figure 1 shows the process that we have applied for causal extraction. This causal mining approach starts by separating the Pubmed abstracts into sentences and applying syntactic and semantic tagging using POS tagger and a semantic tagging mechanism. Actors in the sentences are identified using CRF and based on the identified entities; the sentences are classified into causal or non-causal.

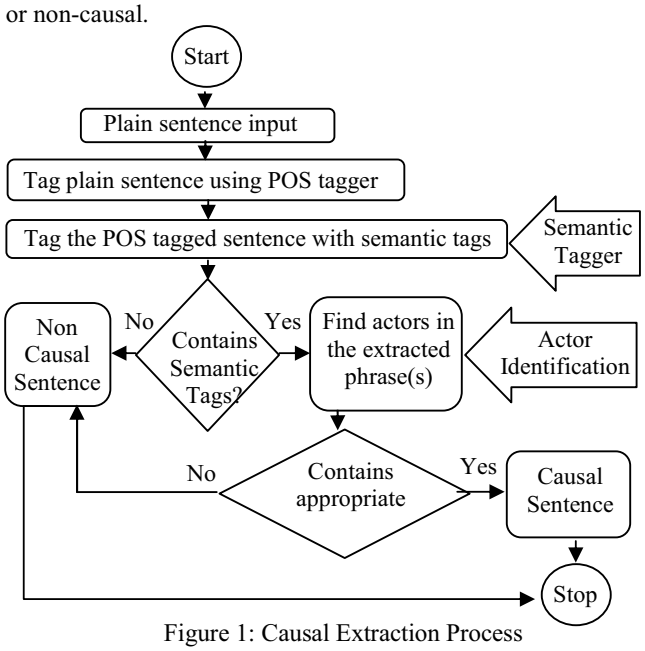

Figure 1: Causal Extraction Process

Building a Keyterm Dictionary

After analysis of the sentences that were reviewed by the domain experts, it was found that each causal sentence contains one or more causal phrases or terms. For example, *"In the elderly, systolic blood pressure increases because of arterial stiffness produced by structural alteration of arterial wall occurring with aging."* shows the relation between "*systolic blood pressure*" and "*arterial stiffness*" using the phrase "*increases because of*". These relations are mainly defined by the presence of such key-phrases (or keyterms) and sentence. In some cases, the presence of relational words and the keywords may not mean that the sentence is causal. For example, in sentence *"Numerous treatable causes of anorexia and weight loss exist."* even though the causal term "*causes*" is present, the sentence does not convey causal behavior. The relational words do not always appear as a keywords or key-phrases. Sentences that do not contain such a relationship are termed Non-Causal. For example, in sentence *"A small number of preventive services are recommended for all adults, ages 65 years and older."* no keyterms are present to show causality.

Detection of the keywords is a Named Entity Recognition (NER) task. NER is a technique that finds the token boundary and the semantic category for particular terms occurring in the text. There are different approaches to NER. We used a dictionary approach to identify the keywords/key-phrases based on the review of a domain expert. Once the terms or phrases are extracted, a keyterm dictionary is constructed. A total of 207 keywords were extracted from the abstracts. Table 2 shows some of the keyterms.

Table 2. Some Keyterms extracted from geriatric abstracts

associated	cause	contribute	reduce	effect
association	causing	contributing	inhibited	affect
correlated	due	contribution	decrease	impact
correlation	because	contributors	degrade	predict

POS Tag triplets

The simplest approach is to find the occurrence of the keyword in the sentence using simple string matching algorithm or regular expressions. In this way, all possible forms of the keyword can be extracted. Another method is to apply a syntactic tag to the sentence and detect a syntactic tag sequence along with the keyword. This restricts the detection of the keyword only if the keyword occurs in a certain form thus reducing noise. This approach involves the steps given below.

1. Apply POS tagging to the sentence.
2. Search for causal keyterms in the POS tagged sentence.
3. Once a match is found, extract the causal term, one term before the causal term (pre word) and one term after the causal term (post word) along with their POS tags from the tagged sentence.
4. Extract the three POS tags from the phrase in step 3.
5. This forms the POS tag triplet.

After extracting the POS tag triplets from the sentences, a mapping of the POS tags, shown in Figure 2, was formed.

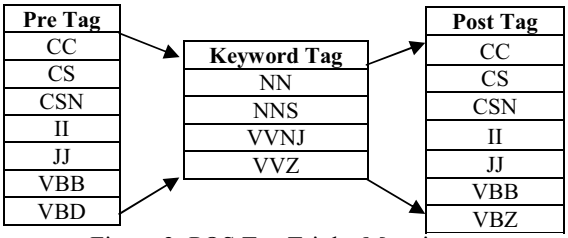

Figure 2: POS Tag Triplet Mapping

Causal Keyterms and Semantic Tag Table Construction

The keyterm table shown in Table 2 was used as the startup dictionary for constructing the semantic tag table. As more care categories were added as part of the research, more causal keyterms were discovered to form a new keyterm dictionary. The new dictionary has a comprehensive set of causal keywords that is used for causal extraction.

Semantic Groups

To get a better understanding of the nature of the sentences that are extracted as causal, the causal keywords were arranged into groups with a name or tag assigned to each group. This approach was applied to the keyword dictionary and the causal

keyterms were divided into 9 different groups. Table 3 shows the 9 semantic groups and the tags assigned to each of these groups. Semantic tagging makes information discovery as efficient meaningful. The process for semantic tagging is given below.

1. POS tagging of the applied on the sentence.
2. Extract only the POS tag sequence from the sentences.
3. Search the POS tag sequence for the POS tag triplet.
4. If a POS tag sequence is found, extract the corresponding text phrase from the sentence.
5. Search the text phrase and check if it contains a causal keyword.
6. If a causal keyword is found, replace the causal keyword with the semantic tag.
7. Store the sentence with the semantic tag back into the sentence set.

TABLE 3: SEMANTIC GROUPS (SHORT LIST)

TASO	TCAU	TCON	TDEC	TEFF
associated	cause	contribute	reduce	effect
connected	create	facilitate	inhibited	affect
correlated	due	account	decrease	impact
correlation	factor	plays	degrade	predict
TINC	**TIND**	**TRES**	**TOTH**	
exacerbate	indicative	result	carries	
exasperate	indicator	resulting	experience	
enhance	identify	results	incidence	
enhancing	tended	resulted	problem	

Identifying Actors in Sentences

Actor is a term used to indicate presence of specific entity in a text. This entity is either the "cause" or the "effect" in the causal sentence. The geriatric care domain contains several actors that are specific to geriatric care literature. For example in the sentence "*Stress causes heart attacks."* the actors are "*Stress",* and *"heart attacks"*. These actors represent the cause and effect that forms a causal sentence. We use CRF as a learning model for identifying these actors.

Creating Training Data

The CRF model learns on the features of the words in a sentence. We use the CRF method provided by the Mallet package built on the Java platform for statistical natural language processing, classification, clustering, topic modeling, information extraction, and other machine learning applications. This model accepts training data in the format "*Word feature1 feature2 feature3…….feature N Label*". Table 4 shows an example of a sentence in the training set.

TABLE 4: SAMPLE CRF TRAINING DATA

Sentence	POS Tags	Shallow Parser Tags	Actor/ Non-Actor Labels
Stress	NN	B-NP	Actor
causes	VVZ	I-NP	O
heart	NN	I-NP	Actor
attacks	NN	I-NP	Actor
.	.	O	O

We manually annotated 800 causal sentences that were chosen across all geriatric care-categories to avoid any biasing. The 800 sentences are purely causal in nature and contain actors that very well defined. Once the model was trained, we tested it on the causal sentences extracted after semantic tagging.

IV. EXPERIMENTS AND RESULTS

In the process of implementing the system, several experiments were conducted at every stage. The experiments were run on 42 care categories and a total of 19725 sentences to determine the performance of causal extraction after the implementation of individual modules to the research work. These sentences were manually annotated by two domain experts with a 90% agreement.

Calculation of results

For evaluating the results we computed the precision, recall, false positive rate, f-score and accuracy [25].

Experiments on Applying Semantic Tags

Once the sentences were semantically tagged, as per the first part of the causal extraction process in Figure 1, if a sentence contains a semantic tag, it is marked, causal; if not then it is marked non-causal. For validation, a new set of sentence was identified that contained unknown abstracts from the geriatric domain. This set contained 164 sentences and was manually classified by the domains experts. The tests performed on the validation set showed an improvement to the results found by the semantic tagging. The results are given in Table 5.

TABLE 5: PERFORMANCE OF SEMANTIC TAGGING ON VALIDATION SET

Sentences	TP	TN	FN	FP
164	42	99	10	13
F-Measure	**Precision**	**Recall**	**FPR**	**Accuracy**
78.50%	76.36%	80.77%	11.60%	85.98%

Experiments on Actor Identification

Actor identification was performed using CRF with the Mallet tool. This process involved creating a training data and generating a trained model to test the test set. Actor identification was performed only on those sentences that were marked as causal at the end of the semantic tagging procedure. The reason for doing this is that the aim is to identify causal sentences and only those sentences that contain a semantic tag can indicate causal behavior and hence can be used to identify actors. Once actor(s) are identified in a sentence, as per the final step of the causal extraction process in Figure 1, only those sentences that contain actor(s) are marked as causal; if not then they are marked non-causal. The result in Table 6 shows an increase in precision and overall accuracy and an improved false positive rate. The remaining 10 sentences moved into the non-causal category.

TABLE 6: PERFORMANCE OF SEMANTIC TAGGING AND ACTOR IDENTIFICATION ON VALIDATION SET

Sentences	TP	TN	FN	FP
154	37	99	9	9
F-Measure	**Precision**	**Recall**	**FPR**	**Accuracy**
80.43%	80.43%	80.43%	8.33%	88.31%

Performance of system on all Geriatric domains

Once the tests were performed on the validation set, it was partly confirmed that the system was capable of extracting causal sentences. The confirmation of the tests can be achieved only after executing the system on the all the care-categories and comparing the results across them. The results after executing the system on all the 42 care-categories are given in Figure 3.

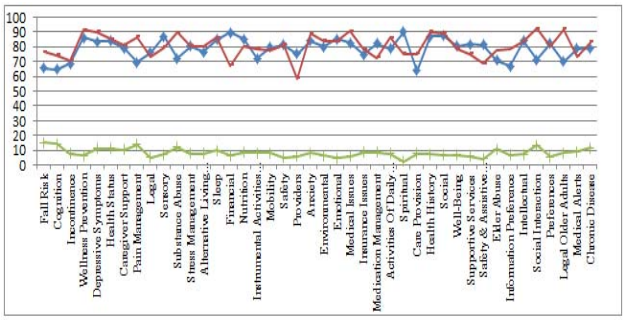

Figure 3: Performance on all geriatric domains

V. CONCLUSION

The causal extraction system works in a multi-layered fashion combining semantic tagging, dictionaries and machine learning approaches to obtain pure causal sentences. The complex structure of the sentences across all the geriatric domains makes the task a lot harder. Although the system provides an overall precision of 79.54%, recall of 81% and an accuracy of 89%, the false positive rate of the system is at 8% which was identified to be a problem associated with sentences that contain multi-causal associations and co-referencing issues due to which actor identification is unable to discover the actors.

FUNDING

This project was supported by a grant Automatic Acquisition of Knowledge for Senior Care Planning from the National Institute of Health (NIH) Small Business Innovation Research (SBIR) grant number 7R44AG038207-03.

REFERENCES

[1] Inouye, Sharon K., et al. "Geriatric syndromes: clinical, research, and policy implications of a core geriatric concept." *Journal of the American Geriatrics Society* 55.5 (2007): 780-791.H. Jiawei and M. Kamber. Data mining: concepts and techniques. San Francisco, CA, itd: Morgan Kaufmann, 5, 2001.

[2] S. Theodoridis and K. Koutroumbas. Pattern recognition. 2006.

[3] S.M. Weiss. Text mining: predictive methods for analyzing unstructured information. Springer-Verlag New York Inc, 2005.

[4] G. Dreyfus. Neural networks: methodology and applications. Springer-Verlag New York Inc, 2005.

[5] K. Gurney and K.N. Gurney. An introduction to neural networks. CRC Press, 1997.

[6] I. Steinwart and A. Christmann. Support vector machines. Springer Verlag, 2008.

[7] US National Library of Medicine National Institutes of Health.

[8] G.G. Chowdhury. Natural language processing. Annual review of information science and technology, 37(1):51-89, 2003.

[9] H. Van Halteren. Syntactic wordclass tagging, volume 9. Springer, 1999.

[10] E. Brill. Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. Computational linguistics, 21(4):543-565, 1995.

[11] K. Toutanova, D. Klein, C.D. Manning, and Y. Singer. Feature-rich part-of-speech tagging with a cyclic dependency network. In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1, pages 173-180. Association for Computational Linguistics, 2003.71

[12] L. Smith, T. Rindesch, W.J. Wilbur, et al. Medpost: a part-of-speech tagger for biomedical text. Bioinformatics, 20(14):2320-2321, 2004.

[13] C.S.G. Khoo, J. Kornfilt, R.N. Oddy, and S.H. Myaeng. Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. Literary and Linguistic Computing, 13(4):177-186, 1998.

[14] D. Marcu and A. Echihabi. An unsupervised approach to recognizing discourse relations. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pages 368-375. Association for Computational Linguistics, 2002.

[15] R. Girju and D. Moldovan. Text mining for causal relations. In Proceedings of the FLAIRS Conference, pages 360-364. AAAI Press, 2002.

[16] G.A. Miller. Wordnet: a lexical database for english. Communications of the ACM, 38(11):39-41, 1995.

[17] C.F. Meyer. English corpus linguistics: An introduction. Cambridge Univ Pr, 2002.

[18] Y. Wilks and M. Stevenson. Sense tagging: Semantic tagging with a lexicon. In Proceedings of the SIGLEX Workshop Tagging Text with Lexical Semantics: What, why and how, pages 47-51, 1997.

[19] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[20] C. Sutton and A. McCallum. An introduction to conditional random fields for relational learning. Introduction to statistical relational learning. MIT Press, 2006. 72

[21] S.D. Sudarsan. Signal Detection Framework Using Semantic Text Mining Techniques. ProQuest LLC. 789 East Eisenhower Parkway, PO Box 1346, Ann Arbor, MI 48106. Tel: 800-521-0600; Web site: http://www. proquest. com/en-US/products/dissertations/individuals. shtml, 2009.

[22] A.K. McCallum. Mallet: A machine learning for language toolkit. 2002.

[23] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2):257-286, 1989.

[24] Mehrabi, S., Krishnan, A., Tinsley, E., Sligh, J., Crohn, N., Bush, H., ... & Palakal, M. (2013, December). Event Causality Identification Using Conditional Random Field in Geriatric Care Domain. In Machine Learning and Applications (ICMLA), 2013 12th International Conference on (Vol. 1, pp. 339-343). IEEE..

[25] I.H. Witten, E. Frank, and M.A. Hall. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2011.