



Blending **Conceptual** and **Evolutionary** Couplings to Support Change Impact Analysis in Source Code

Huzefa Kagdi Winston-Salem State, NC

Malcom Gethers and Denys Poshyvanyk The College of William and Mary, VA

Michael Collard The University of Akron, OH

Research Goal

- Develop a new and improved Impact Analysis (IA) approach by combining existing solutions
 - Information Retrieval (IR), Latent Semantic Indexing (LSI)
 - Mining Software Repositories (MSR), Itemset Mining
- Employ single + multiple version analysis

```
/**
 * Append a new page, creating followup frames (but not headers/footers),
 * and return the page number.
 * @param masterPageName the name of the master page to use for this new page.
 */
KWPage appendPage(const QString &masterPageName = QString());
.....
```

```
r472031 | bmeyer | 2005-10-19 11:14:47 -0400 (Wed, 19 Oct 2005) |
1 line
```

Changed paths:

```
 M   /trunk/KDE/kdegames/ktron/ktron.cpp
 M   /trunk/KDE/kdegames/ktron/main.cpp
```

move scope

Impact Analysis in Source Code

Fix a bug



```
// reallocate,
void Vector::reallocate(unsigned newsize, bool copy)
{
    // resize to the given size
    double* temptr = new double[newsize];
    // ...
}
```

Estimates



Approach

{

```
int main(int argc, char **argv)
{
    KAboutData about("helloworld",
        KAboutData::
        about.addAuthor( "Huzefa Kag",
        KCmdLineArgs::init(argc, arg
        KCmdLineArgs::addCmdLineOpti
        KApplication app;
        HelloWorld *mainWin = 0;

    if (app.isRestored())
    {
        RESTORE(HelloWorld);
    }
    else
    {
```

```
int mai
{
    KAB
    KAboutDa
    abo
    about.ad
    KCm
    KCm
    KCm
    KAp
    KAppl
    Hel
```

```
int main(int argc, char **argv)
{
    KAboutData about("helloworld",
        KAboutData:
        about.addAuthor( "Huzefa Kag",
        KCmdLineArgs::init(argc, arg
        KCmdLineArgs::addCmdLineOpti
        KApplication app;
        HelloWorld *mainWin = 0;

    if (app.isRestored())
    {
        RESTORE(HelloWorld);
    }
    else
    {
```

```
int main(int argc, char **argv)
{
    KAB
    a
    KCm
    KCm
    KAp
    KAp
    H
```

```
int main(int argc, char **argv)
{
    KAboutData about("helloworld",
        KAboutData:
        about.addAuthor( "Huzefa Kag",
        KCmdLineArgs::init(argc, arg
        KCmdLineArgs::addCmdLineOpti
        KApplication app;
        HelloWorld *mainWin = 0;

    if (app.isRestored())
    {
        RESTORE(HelloWorld);
    }
    else
    {
```

}

Impact Set

Combined Approach for IA

- **Step 1: Select** the first software entity for which IA needs to be performed
- **Step 2: Compute** conceptual couplings with IR methods from the release (source code comments and identifiers) of a software system in which the first entity is selected
- **Step 3: Mine** evolutionary couplings from the source code change history (commits in the source code repository), occurred before the release in Step 2
- **Step 4: Compute** the estimated impact set from the disjunctive combination (union) of couplings computed in steps 3 and 4

Motivating Example

- In *Apache httpd* commit# 888310 addresses the bug#47087
- Three source code files were changed
 - /modules/http/http_filters.c
 - /modules/http/http_protocol.c
 - /server/protocol.c

Rank	Conceptual	Evolutionary	Combined
1	/server/protocol.c	/modules/http/byterange_filter.c	/server/protocol.c
2	/modules/proxy/mod_proxy_http.c	/modules/http/http_protocol.c	/modules/proxy/mod_proxy_http.c
3	/modules/debugging/mod_bucketeer.c	/modules/proxy/mod_proxy_ftp.c	/modules/http/byterange_filter.c
4	/server/core_filters.c	/server/core.c	/modules/http/http_protocol.c
5	/modules/http/byterange_filter.c	/include/ap_mmn.h	/server/core_filters.c

Extracting Semantic Information with LSI

- Source code (release) -> Corpus (doc = method)
- Preprocessing: split_identifiers & splitIdentifiers
- Vector space = term-by-document matrix
- Singular Value Decomposition -> LSI subspace

	document	shape	set	frame	view	
KWDocument::addShape	1	10	6	15	5	...
KWDocument::removeShape	1	7	5	16	6	...

Computing Conceptual Similarity

Cosine between document vectors [Poshyvanyk, Marcus]

	KWDocument::addShape	KWDocument::removeShape	KWDocument::insertPage	
KWDocument::addShape	1	0.78	0.24	...
KWDocument::removeShape	0.78	1	0.27	...
KWDocument::insertPage	0.24	0.27	1	...

```
void KWDocument::addShape(KoShape *shape)
{
    // KWord adds a couple of dialogs (like KWFrameDialog) which will not call addShape(), but
    // will call addFrameSet(). Which will itself call addFrame()
    // any call coming in here is due to the undo/redo framework or for nested frames

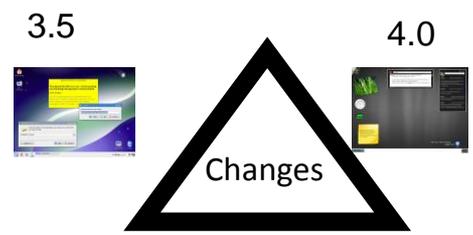
    KWFrame *frame = dynamic_cast<KWFrame*>(shape->applicationData());
    if (frame == 0) {
        KWFrameSet *fs = new KWFrameSet();
        fs->setName(shape->shapeId());
        frame = new KWFrame(shape fs);
    }
    Q_ASSERT(frame->frameSet());
    addFrameSet(frame->frameSet());

    foreach (KoView *view, views()) {
        KWCanvas *canvas = static_cast<KWView*>(view)->kwcanvas();
        canvas->shapeManager()->add(shape);
    }
}
```

⇒
0.78

```
void KWDocument::removeShape(KoShape *shape)
{
    KWFrame *frame = dynamic_cast<KWFrame*>(shape->applicationData());
    if (frame) { // not all shapes have to have a frame. Only top-level ones do.
        KWFrameSet *fs = frame->frameSet();
        Q_ASSERT(fs);
        if (fs->frameCount() == 1) // last frame on FrameSet
            removeFrameSet(fs); // frame and frameset will be deleted when the shape is deleted
        else
            fs->removeFrame(frame);
    } else { // not a frame, but we still have to remove it from views.
        foreach (KoView *view, views()) {
            KWCanvas *canvas = static_cast<KWView*>(view)->kwcanvas();
            canvas->shapeManager()->remove(shape);
        }
    }
}
```

Development Versions



Fix a bug



δ_i

=



δ_k

δ_n



=

```

{
int main(int argc, char **argv)
{
    KAboutData about(
        about.addAuthor(
            KCmdLineArgs::in
            KCmdLineArgs::ad
            KApplication app
            HelloWorld *main

    if (app.isRestor
    {
        RESTORE(Hell
    }
    else
    {
int main(int argc, char **argv)
{
    KAboutData about("helloworld
        KAboutData:
        about.addAuthor( "Huzefa Kad
            KCmdLineArgs::init(argc, arg
            KCmdLineArgs::addCmdLineOpti
            KApplication app;
            HelloWorld *mainWin = 0;

    if (app.isRestored())
    {
        RESTORE(HelloWorld);
    }
    else
    {
}
}
}
}
    
```

```

r472031 | bmeyer | 2005-10-19
11:14:47 -0400 (Wed, 19 Oct
2005) | 1 line

Changed paths:
M
/trunk/KDE/kdegames/ktrn/ktr
on.cpp
M
/trunk/KDE/kdegames/ktrn/mai
n.cpp
    
```

Subversion Commit History

r472031 | bmeyer | 2005-10-19 11:14:47 -0400 (Wed, 19 Oct 2005) | 1 line

Changed paths:

- M /trunk/KDE/kdegames/ktron/ktron.cpp
- M /trunk/KDE/kdegames/ktron/main.cpp

move scope

r472031 | bmeyer | 2005-10-19 11:14:47 -0400 (Wed, 19 Oct 2005) | 1 line

Changed paths:

- M /trunk/KDE/kdegames/ktron/ktron.cpp
- M /trunk/KDE/kdegames/ktron/main.cpp

move scope

r472031 | bmeyer | 2005-10-19 11:14:47 -0400 (Wed, 19 Oct 2005) | 1 line

Changed paths:

- M /trunk/KDE/kdegames/ktron/ktron.cpp
- M /trunk/KDE/kdegames/ktron/main.cpp

move scope

r472031 | bmeyer | 2005-10-19 11:14:47 -0400 (Wed, 19 Oct 2005) | 1 line

Changed paths:

- M /trunk/KDE/kdegames/ktron/ktron.cpp
- M /trunk/KDE/kdegames/ktron/main.cpp

move scope

r472031 | bmeyer | 2005-10-19 11:14:47 -0400 (Wed, 19 Oct 2005) | 1 line

Changed paths:

- M /trunk/KDE/kdegames/ktron/ktron.cpp
- M /trunk/KDE/kdegames/ktron/main.cpp

move scope

r472031 | bmeyer | 2005-10-19 11:14:47 -0400 (Wed, 19 Oct 2005) | 1 line

Changed paths:

- M /trunk/KDE/kdegames/ktron/ktron.cpp
- M /trunk/KDE/kdegames/ktron/main.cpp

move scope

r472031 | bmeyer | 2005-10-19 11:14:47 -0400 (Wed, 19 Oct 2005) | 1 line

Changed paths:

- M /trunk/KDE/kdegames/ktron/ktron.cpp
- M /trunk/KDE/kdegames/ktron/main.cpp

move scope

r472031 | bmeyer | 2005-10-19 11:14:47 -0400 (Wed, 19 Oct 2005) | 1 line

Changed paths:

- M /trunk/KDE/kdegames/ktron/ktron.cpp
- M /trunk/KDE/kdegames/ktron/main.cpp

move scope

r472031 | bmeyer | 2005-10-19 11:14:47 -0400 (Wed, 19 Oct 2005) | 1 line

Changed paths:

- M /trunk/KDE/kdegames/ktron/ktron.cpp
- M /trunk/KDE/kdegames/ktron/main.cpp

move scope

Evolutionary Couplings

Software entities that typically co-change in the version archives [Gall'98, Zimmermann'04]

	Button.java	Events.java	GUI.java	Filter.java
Button.java	10	0	3	0
Events.java	0	2	0	0
GUI.java	3	0	4	0
Filter.java	0	0	0	5

Evolutionary Pattern

{Button.java, GUI.java}

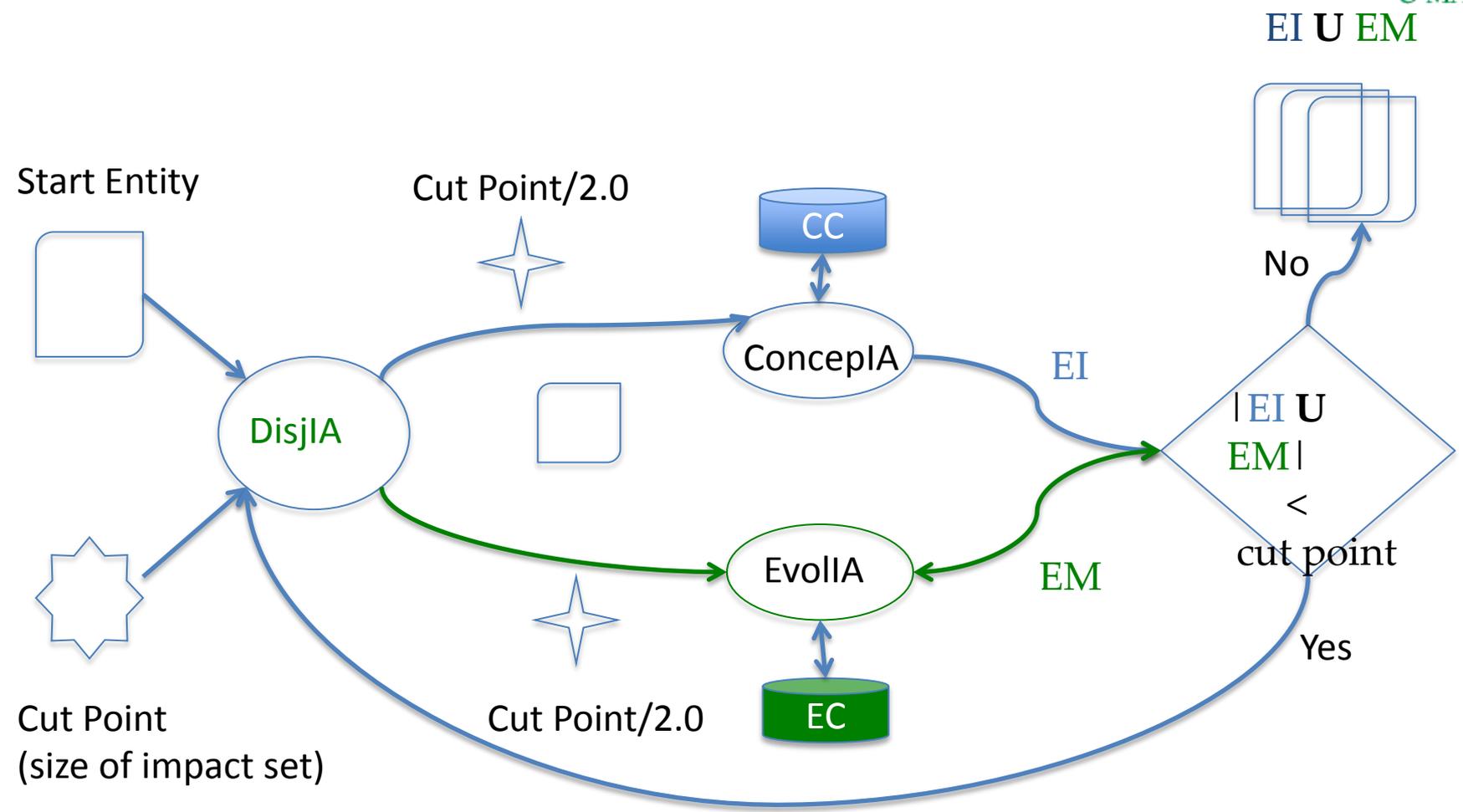
co-changed in 3 commits (support value)

with itemset mining

Change prediction rules from association rules

{GUI.java} \Rightarrow {Button.java, GUI.java} with a 75% probability, i.e., confidence value

Procedure for Computing Impact Set



Conceptual Couplings (CC) are ranked by cosine similarity values
 Evolutionary Couplings (EC) are ranked by confidence and support values

Case Study

- Four open source software systems, namely *Apache httpd*, *ArgoUML*, *iBatis*, and *KOffice*
- Widely used metrics precision (false positives) and recall (false negatives) used for accuracy measure
- Commits in the version history used as ground truth, i.e., actual impact sets
- Accuracy assessment performed for various estimate impact set sizes

Research Questions

- **RQ1:** Does combining conceptual and evolutionary couplings improve the accuracy of IA when compared to the two standalone techniques?
- **RQ2:** Does the choice of granularity, *i.e.*, file or method, affect the accuracy of IA of standalone techniques and their combination?

Evaluation Dataset

System	Release History	LSI Indexed Release	# of Commits Train (test)	# of Entities Train (test)
Apache (httpd)	2.2.9-2.3.5	2.2.3	1736 (287)	2086 (982)
ArgoUML	0.24-0.28	0.28	3375 (773)	4217 (621)
iBatis	3.0.0-190_b1 - 3.0.0-240_b10	3.0.0-216	108 (40)	461 (118)
KOffice 2.0.91	2.0.0-2.0.91	2.0.91	2749 (522)	5580 (1072)
KOffice 2.0.1	2.0.0-2.0.2	2.0.1	763 (255)	1233 (533)
KOffice 2.0.1*	2.0.0-2.0.2	2.0.1	577 (192)	5530 (1438)

* Denotes method level granularity processed with *srcDiff* [Collard'03]

Null Hypotheses

- $H_{0_{CP}}$: Combining conceptual and evolutionary couplings *does not significantly* improve precision results of impact analysis compared to conceptual couplings
- $H_{0_{CR}}$: Combining conceptual and evolutionary couplings does not significantly improve recall results of impact analysis compared to conceptual couplings
- $H_{0_{EP}}$: Combining conceptual and evolutionary couplings does not significantly improve precision results of impact analysis compared to evolutionary couplings
- $H_{0_{ER}}$: Combining conceptual and evolutionary couplings does not significantly improve recall results of impact analysis compared to evolutionary couplings

Results of Wilcoxon Signed-Rank Test

System	Granularity	$H_{0\ CP}$	$H_{0\ CR}$	$H_{0\ EP}$	$H_{0\ ER}$	Null Hypothesis
Apache (httpd)	File	0.0002	0.0003	0.0001	0.0003	Rejected
ArgoUML		0.0050	0.0039	< 0.0001	< 0.0001	Rejected
iBatis		0.0126	0.0126	0.0001	0.0002	Rejected
KOffice 2.0.91		< 0.0001	< 0.0001	< 0.0001	< 0.0001	Rejected
KOffice 2.0.1		< 0.0001	< 0.0001	< 0.0001	< 0.0001	Rejected
KOffice 2.0.1*	Method	< 0.0001	< 0.0001	< 0.0001	< 0.0001	Rejected

In all cases considered for our dataset we obtained a p-value less than 0.05, indicating that the improvement in accuracy obtained is not by chance

Threats to Validity

- Commits used as gold standard for accuracy computation
 - Not all the entities in a commit maybe related to a single change request
 - All the entities related to a single change request maybe present in a single commit
 - Developer established actual change-sets
- Granularity levels of file and method
- Statistically significant results for the four open source system may not generalize

Related Work

- Various dependency-analysis methods are already investigated in the literature
 - Call graphs, program slicing, hidden dependency analysis [Chen, Rajlich, Yu], Lightweight static analysis approaches [Moonen], concept analysis [Tonella], dynamic analysis [Law], hypertext systems, documentation systems, UML models [Briand], and Information Retrieval [Antoniol]
 - Coupling measures have been also used to support impact analysis in OO systems [Briand, Wilkie]
 - Comparison of different impact analysis algorithms [Osro]
- An IR (single version) and MSR (multiple version) combination has not been investigated for IA previously

Conclusions

- Combining conceptual and evolutionary couplings does improve accuracy of IA
- Recall improvements of up to 20% over the conceptual technique in *KOffice* and up to 45% over the evolutionary technique in *iBatis*
- Varying granularity levels does impact accuracy of individual methods; however, combining conceptual and evolutionary couplings maintains the accuracy gains

Future Work

- Devise and empirically validate other combinations (e.g., weighed contributions of entities from each coupling based on the amount of change history considered)
- Include static and dynamic analysis information, and application of IR on multi-version artifacts (e.g., commit messages and bug reports)
- Provide IA support beginning from a high-level textual change request
- Comparative studies with other approaches (e.g., structural metrics)

- Work supported by NSF CCF-1016868 and NSF CCF-1016887 grants



- Questions?